Contextuality as a resource for qubit quantum computation

Robert Raussendorf[1], Dan E. Browne[2], Nicolas Delfosse[3,4,5], Cihan Okay[6], Juan Bermejo-Vega[7,8]

1: Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada,
2: Department of Physics and Astronomy, University College London, Gower Street, London, UK,
3: Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, Canada,
4: IQIM, California Institute of Technology, Pasadena, CA, USA,
5: Department of Physics and Astronomy, University of California, Riverside, California, 92521, USA,
6: Department of Mathematics, University of Western Ontario, London, Ontario, Canada,
7: Max-Planck Institut für Quantum Optics, Theory Division, Garching, Germany,
8: Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin, Germany

(Dated: November 30, 2015)

We describe a scheme of quantum computation with magic states on qubits for which contextuality is a necessary resource possessed by the magic states. More generally, we establish contextuality as a necessary resource for all schemes of quantum computation with magic states on qubits that satisfy three simple postulates. Furthermore, we identify stringent consistency conditions on such computational schemes, revealing the general structure by which negativity of Wigner functions, hardness of classical simulation of the computation, and contextuality are connected.

PACS numbers: 03.67.Mn, 03.65.Ud, 03.67.Ac

I. INTRODUCTION

Contextuality [1] - [5] has recently been established as a necessary resource for quantum computation by injection of magic states (QCSI). This was first achieved for the case of qudits [6], where the Hilbert space dimension of the local systems is an odd prime or a power of an odd prime, and subsequently for the case of rebits [7], where the Hilbert space dimension of the local systems is 2, but the density matrix is constrained to be real.

The scheme of QCSI [8] deviates from the standard circuit model in that the allowed state preparations, unitary transformations and measurements are restricted to non-universal and, in fact, efficiently classically simulable operations. Computational universality is restored by the capability to inject so-called magic states. The source of computational power thus shifts from the gates to the magic states.

Before the analysis of the magic states as resources can begin, it needs to be clarified in which sense the restricted state-preparations, unitaries and measurements available in QCSI are not quantum resources. These operations are certainly not entirely classical. For example, highly entangled states can be created by them. The near-classicality of these operations is explained in terms of a Wigner function; See [6], [7], [9] - [11].

Wigner functions [12] - [15] describe quantum states in phase space. They are quasi-probability distributions, and as such the closest quantum analogue to joint probability distributions of position and momentum in classical statistical mechanics. The difference is that Wigner functions can take negative values, and this negativity is a signature of quantumness [16], [17].

For QCSI on qudits or rebits, Wigner functions provide a computational notion of classicality [9], [15], [7]. Namely, if the initial quantum state has a non-negative Wigner function, then the entire quantum computation can be efficiently classically simulated. Wigner function negativity is thus necessary for quantum speedup.

After the roles of Wigner function negativity and contextuality have been clarified for qudits and rebits, in this paper we investigate them for the yet unresolved case of qubits. The case of local dimension 2, into which the rebit case forays, is complicated by the fact that the Wigner function for infinite dimension [12] cannot be adapted to it [14], [15], [18], [19], by the presence of state-independent contextuality with Pauli observables [3], and Bell inequalities based on stabilizer operators [20]-[22].

We impose the following three constraints on the QCSI schemes we discuss: (P1) *The computational scheme is tomographically complete.* That is, with the available operations the density matrix ρ of any n-qubit quantum state can be fully measured, and that (P2) *The Wigner function describing the computational scheme is informationally complete*, i.e., any n-qubit quantum state ρ can be unambiguously reconstructed from its Wigner function W_ρ. Finally, (P3) *The measurements available in QCSI must not introduce negativity into the Wigner function of the processed quantum state.*

Requirement (P3) is the very basis for the usefulness of Wigner functions in the description of QCSI, namely to reveal the near-classicality of QCSI without the magic states. It is certainly in line with the approach taken for qudits and rebits. However, (P3) is trickier than might at first appear. For a start, we do not require a counterpart of (P3) for the unitary operations available in QCSI, and imposing it would indeed be too restrictive. Those unitaries may introduce large amounts of negativity into the Wigner function without compromising efficient classical simulability.

In this paper, we provide a common structural framework for QCSI schemes on qubits which satisfy the above postulates (P1) - (P3), and demonstrate that there are such QCSI schemes. For those, we establish contextuality in the magic states as a necessary quantum resource.

II. RESULTS AND OUTLINE

A. Summary of results

Our main results are the following:

1. For all QCSI schemes on qubits satisfying the postulates (P1)-(P3), for $n \geq 2$ qubits, contextuality is necessary for quantum computational universality (Theorem 3) and for speedup (Theorem 4).

2. There is at least one family of QCSI schemes which satisfies the postulates (P1)-(P3).

3. For qubits, two notions of classicality in QCSI agree, namely the notion based on the existence of a non-contextual HVM and the notion based on efficient classical simulation by sampling (Theorem 4).

4. For qubits, the unitary gates allowed in QCSI do in general not preserve positivity of Wigner functions and do not transform Wigner functions covariantly. This does not affect efficient classical simulability.

5. As for qudits, the Wigner function is a critical tool for endowing the operations of QCSI (not invoking magic states) with a notion of near-classicality.

The last three points require explanation. To begin, we observe that three notions of classicality are considered in the literature to describe the limitations of QCSI without magic states, namely (i) non-contextuality, (ii) efficient classical simulation by sampling from a non-negative Wigner function [9], and (iii) efficient classical simulation via the stabilizer formalism [23].

Regarding point 3, for qudits in odd prime (power) dimension, the first two of these notions turn out to be the same [6], and the third notion is strictly included [9]. There is thus a robust notion of classicality in QCSI.

For qu*bits*, the situation is more complicated. For example, the phenomenon of state-independent contextuality w.r.t. Pauli observables [3] arises, which is not present in qudits [24]. Also, classical simulability by sampling from a Wigner function is a more restricted notion of classicality than the existence of a non-contextual HVM.

To close the gap between those two notions of classicality, in Section V E we describe a general sampling algorithm which is based on an HVM rather than a non-negative Wigner function. This algorithm has the same range of applicability as the non-contextual HVMs themselves. We thus find that the fundamental classical object, both from the perspective of non-contextuality and from the perspective of efficient classical simulation by sampling, is the non-contextual HVM, and not a positive Wigner function.

Regarding point 4, the situation is in stark contrast to the previously considered cases of qudits [9] and rebits [7], where the Wigner function in question is transformed covariantly and positivity is preserved. As a consequence, for all QCSI schemes on qubits where positivity of the Wigner function is indeed not preserved, positivity cannot be a sufficient resource for speedup (the question is presently open in the qudit case). After the above-mentioned general simulation algorithm based on an HVM, the failure of the considered Wigner functions to transform covariantly and to preserve positivity under the unitary QCSI-gates deals a second blow to the perceived centrality of Wigner functions for the description of QCSI [10], [11], [9], [7].

Regarding point 5, the above limitations notwithstanding, Wigner functions hold up as an organizing principle for near-classicality in QCSI. Specifically, the critical postulate (P3) is formulated in terms of a Wigner function, and this formulation remains adequate. That is, the Wigner function imposes the same constraints on the corresponding QCSI scheme as a more general non-contextual HVM. How can this be?

The answer to this question is that if an input state ρ can be described in terms of a non-contextual HVM, then it corresponds to an ensemble $\mathcal{E}_\rho$ of states,

$$\mathcal{E}_\rho = \{(p_i, \rho_i), i \in \mathcal{I}\},$$

such that there are Wigner functions W^{γ_i} with the property that $W^{\gamma_i}_{\rho_i} \geq 0$, for all $i \in \mathcal{I}$. (The set $\mathcal{I}$ will be specified later.) The reason it remains meaningful to formulate the postulate (P3) in terms of Wigner functions is that for all the above W^{γ_i}, $i \in \mathcal{I}$, the constraints placed by (P3) on the QCSI scheme in question are the same. This is explained in detail in Section V F.

Remark: In our results on efficient simulation by sampling (Theorems 1 and 4), we assume the sampling sources as given, and only count the operational cost of processing the samples in the simulation. This assumption holds, for example, when each magic state injected to the computation has support only on a bounded number of qubits [9],[7]. However, there is strong indication that probability distributions exist which can be efficiently prepared by quantum means but are hard to sample from classically [25] - [30]. In view of those, Theorems 1 and 4 specify the computational cost of classical simulation relative to a sampling source, similar to the complexity of an algorithm relative to an oracle.

B. Relation to previous work

The role of positive Wigner functions for QCSI has previously been discussed in [10], [11], [9] and [7]. Of those works, [10], [11] and [7] address 2-level systems. In [10] and [11], multiple Wigner functions are considered simultaneously, and for (near-) classicality it is required that the processed quantum states are positive w.r.t. *all* those Wigner functions. This requirement severely limits the scope of the free operations of QCSI. By contrast, in our approach a sufficient requirement for near-classicality is that the initial state is positive w.r.t. a *single* Wigner function, and this requirement is relaxed even further (see the discussion in Section II A, and Section V F).

From the perspective of Wigner functions, the present work is an extension of [9] an [7]. In [9], systems of qudits in odd prime power dimension are discussed. While [7] addresses 2-level systems, the density matrices therein are constrained to be real. Here we lift that restriction.

The present work differs from all above works in one critical respect. Namely, in [10], [11], [9] and [7], positivity of the considered Wigner function is preserved under all operations of QCSI which do not invoke magic states. For the present discussion of qubits, this is not the case. Positivity of the Wigner function remains preserved under the measurements available in QCSI, but not necessarily under the unitaries.

From the perspective of contextuality, the present work is an extension of [6] (qudits of odd prime power dimension) and [7] (rebits). In [6], contextuality was first established as a necessary resource for QCSI. The present work generalizes the approach of operational restrictions previously applied to the rebit case [7]. Here, those operational restrictions derive from the postulate (P3).

C. Outline

This paper is structured as follows. In Sections III - V we analyze the general structure of QCSI schemes defined by the postulates (P1) - (P3), and in Section VI we explicitly construct a QCSI scheme on qubits for which contextuality in the magic states is a necessary quantum mechanical resource. Regarding the former part, in Section III, we work out the implications of the postulates (P1) - (P3) for QCSI schemes. We give a prescription for how to construct QCSI schemes starting from the phase convention γ for the Heisenberg-Weyl operators. Section IV discusses the role of Wigner functions for QCSI. In particular, we present an efficient classical simulation of QCSI for magic states with non-negative Wigner function (Algorithm 1). Section V is on the role of contextuality. We show that state-independent contextuality is absent from all QCSI schemes satisfying the postulates (P1)-(P3), clarify the relation between Wigner function negativity and state-dependent contextuality, and establish the latter as a necessary resource for QCSI with magic states. Finally, we describe an efficient classical simulation algorithm for QCSI for magic states with a non-contextual HVM (Algorithm 2). It contains Algorithm 1 as a special case. We conclude in Section VII.

III. COMPUTATIONAL SETTING AND CONSISTENCY CONDITIONS

In this section we demonstrate that, given the postulates (P1) - (P3), the choice of Wigner function largely determines the corresponding QCSI scheme. In Section III A, we briefly review the model of QCSI. In Section III B, we discuss the general concept of an operational restriction, how it overcomes the phenomenon of state-independent contextuality, and why that is necessary for establishing contextuality of the magic states as a resource for QCSI. In Sections III C - III E, we describe the the compatibility constraints between the constituents of QCSI. In Section III F, we provide an algorithm for constructing the free part of the corresponding QCSI scheme, i.e. the allowed Clifford unitaries, Pauli measurements and stabilizer state preparations.

A. The computational setting

Every QCSI scheme consists of four constituents, namely (i) a set Ω of states that can be prepared within the scheme (the "free" states), (ii) the set $\mathcal{O}$ of observables which can be directly measured, and which in the present discussion always consists solely of Pauli operators, (iii) a group G of unitary gates (the "free gates"), typically taken as the Clifford group or a subgroup thereof, and (iv) the set $\mathcal{M}$ of magic states which render the scheme computationally universal. We thus denote a scheme of QCSI by the quadruple $(\mathcal{O}, G, \Omega, \mathcal{M})$.

The first three of these four constituents are considered "free". The justification for this terminology is that quantum computations built solely from the free operations cannot have a quantum speedup. This near-classicality of the free operations is made precise by an efficient classical simulation algorithm (see Section IV). It states that if the Wigner function of the initial quantum state ρ_{in} can be efficiently sampled from then so can the outcome distribution resulting from evolving ρ_{in} under the free unitary gates and measurements. This simulation result is the very justification for invoking a Wigner function in the description of QCSI.

B. Operational restrictions

When transitioning from local systems of odd prime Hilbert space dimension (qudits) to local systems of Hilbert space dimension 2 (qubits), one encounters a new phenomenon: state-independent contextuality among Pauli-observables [3], [24]. It is incompatible with viewing contextuality as a resource injected into the computation along with the magic states.

The reasons are two-fold. First, within the framework of QCSI, Pauli-measurements are supposed to be free, and if contextuality is already present in those operations, how can it be a resource? Perhaps even worse, for

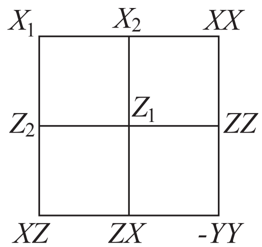


FIG. 1: Mermin's square. For the restriction to CSS-ness preserving operations, the six observables in the top two rows are in the set $\mathcal{O}$ while the observables in the bottom row are in M but not in $\mathcal{O}$. In rebit QCSI they can be measured individually but not jointly. The figure is adapted from [3].

systems of two or more qubits, a contextuality witness can be constructed that classifies *all* quantum states of $n \geq 2$ qubits as contextual [6], including the completely mixed state. Again, how can contextuality be a resource if it is generic?

In this paper, the strategy for coping with state-indepenendent contextuality is to place operational restrictions on the Pauli observables that can be measured in a QCSI scheme. The very concept of QCSI already invokes the notion of an operational restriction, since the operations in QCSI are non-universal by design. Here, additional constraints are placed by the postulate (P3). The rebit case [7] shall serve as a model scenario for the concept of operational restrictions, and we briefly review it for illustration.

Mermin's square embeds into real quantum mechanics (see Fig. 1), and confining to rebit quantum states does therefore not remove state-independent all by itself. Rather, the following operational restriction is put in place. The directly measurable observables are restricted from the set of real Pauli operators to tensor products of Pauli operators Z_i only or X_i only. Accordingly, the free unitaries are restricted from all real Clifford gates to those which preserve the set of Calderbank-Shor-Steane (CSS) stabilizer states [31],[32].

Let us analyze the free measurements, for the case of $n = 2$ rebits, with the CSS restriction. The set $\mathcal{O}$ of directly measurable observables is

$$\mathcal{O} = \{I, Z_1, Z_2, Z_1Z_2, X_1, X_2, X_1X_2\} \times \{\pm 1\}$$

By directly measuring observables from the above set, measurement outcomes of further Pauli observables can be inferred. For example, $X_1Z_2 \notin \mathcal{O}$. Yet, a value for X_1Z_2 can be inferred by measuring the commuting observables X_1 and Z_2 separately, and multiplying the outcomes. Applying this construction to all possible pairs of commuting observables in $\mathcal{O}$, we find the set M of observables whose value can be inferred, namely

$$M = \mathcal{O} \cup \{X_1Z_2, Z_1X_2, Y_1Y_2\} \times \{\pm 1\}.$$

M is thus the set of all real and Hermitian two-qubit Pauli operators. By measurement of observables in the smaller set $\mathcal{O}$ it is thus possible to fully reconstruct all two-rebit density operators.

The next question of interest is which Pauli operators can be measured *jointly*. For example, while both the observables X_1Z_2 and Z_1X_2 are in M and even though they commute, in rebit QCSI they cannot have their values inferred simultaneously. Inferring the value of X_1Z_2 necessitates the physical measurement of the observables X_1 and Z_2, and inferring the value for Z_1X_2 requires the physical measurement of Z_1 and X_2. However, the four observables X_1, Z_2, Z_1 and X_2 do not all commute. The measurement of Z_1 and X_2 to infer the outcome of Z_1X_2 wipes out the value of X_1Z_2, and vice versa.

The fact that the observables X_1Z_2 and Z_1X_2 cannot have their values inferred simultaneously is critical for state-independent contextuality. Namely, the consistency constraint among measurement outcomes for observables in the bottom row of Mermin's square can no longer be experimentally checked, and is thus effectively removed from the square. As a consequence, the remaining available measurements can be described by a non-contextual hidden variable model (HVM). For example, the value assignment $\lambda = 1$ for all observables in Mermin's square becomes consistent. In this way, by imposing an operational restriction, state-independent contextuality disappears from QCSI.

This concludes the review of the rebit case. In the subsequent sections we generalize the notions introduced above and apply them to a wider range of settings. As a final remark, earlier in this section we stated that the operational restrictions must obey certain consistency conditions. The above discussion points to two of them: To give rise to a tomographically complete scheme of QCSI on qubits, the set $\mathcal{O}$ of directly measurable observables must be large enough for the derived set M to comprise all Pauli operators. At the same time, $\mathcal{O}$ must be small enough to dispense with state-independent contextuality.

C. Consistency conditions on G and Ω

We now begin to describe the consistency conditions which must hold between the group G of free unitary gates in QCSI, the set $\mathcal{O}$ of directly measurable observables, and the set Ω of free states. We require that these constituents of QCSI satisfy two constraints, namely

$$g^\dagger O g \in \mathcal{O}, \;\; \forall O \in \mathcal{O}, \forall g \in G, \tag{1}$$

and

$$g|\psi\rangle \in \Omega, \; \forall |\psi\rangle \in \Omega, \forall g \in G. \tag{2}$$

Regarding Eq. (1), if O can be measured, so can $g^\dagger O g$, namely by first applying g, then measuring O and then applying $g^\dagger$. Likewise, if $|\psi\rangle$ can be prepared, so can $g|\psi\rangle$

We regard the set $\mathcal{O}$ of directly measurable observables as primary among the constituents of the free sector of QCSI, and define the group G of free gates and the set Ω

of free states in reference to it. Namely, G is the largest subgroup of the n-qubit Clifford group Cl_n that satisfies the property Eq. (1),

$$G := \{g \in Cl_n |\, g^\dagger O g \in \mathcal{O},\, \forall O \in \mathcal{O}\}. \tag{3}$$

The free states are those that can be prepared by measurement of observables in $\mathcal{O}$. All other states are considered resources, and must be provided externally if needed. That is, $|\psi\rangle \in \Omega$ if and only if there exists an ordered set $\mathcal{O}_{|\psi\rangle} \subset \mathcal{O}$ such that

$$|\psi\rangle\langle\psi| \sim \left(\prod_{O \in \mathcal{O}_{|\psi\rangle}} \left[\frac{I \pm O}{2}\right]\right)(I/2^n). \tag{4}$$

The projectors on the lhs. of Eq. (4) do not necessarily commute. Their temporal order is specified by the ordering in J. The angular brackets denote superoperators. With Eq. (3), $\mathcal{P}_n \subset G$ always holds. Therefore, a totally depolarizing twirl may be implemented, producing $I/2^n$ from any n-qubit state.

The free sector of a QCSI scheme is thus fully specified via Eqs. (3) and (4) by the set $\mathcal{O}$ of directly measurable observables. In Section III E we turn to the question of how $\mathcal{O}$ itself is constructed.

In accordance with the programme outlined in Section I, for the present purpose we enforce the additional requirement that

(P1) The available measurements are tomographically complete for n-qubit states.

D. Wigner functions

A Wigner function is a means of description of QCSIs. The reason for invoking Wigner functions is to characterize the near-classicality of the sector of free operations in QCSI. This proceeds by way of the efficient classical simulation algorithm described in Section IV B.

The Wigner functions considered here are defined on a phase space $V := \mathbb{Z}_2^n \times \mathbb{Z}_2^n$, starting from the Heisenberg-Weyl operators

$$T_{\mathbf{a}} = i^{\gamma(\mathbf{a})} Z(\mathbf{a}_Z) X(\mathbf{a}_X). \tag{5}$$

Therein, $Z(\mathbf{a}_Z) := \bigotimes_{i=1}^n {Z_i}^{\mathbf{a}_{Z,i}}$, $X(\mathbf{a}_X) := \bigotimes_{i=1}^n {X_i}^{\mathbf{a}_{X,i}}$.

The possible phase conventions $\gamma : V \longrightarrow \mathbb{Z}_4$ are constrained only by the requirement that all $T_{\mathbf{a}}$, $\mathbf{a} \in V$, are Hermitian. As we show later, the QCSI schemes considered here and the Wigner functions describing them are both fully specified by γ.

We consider Wigner functions of the form $W_\rho(\mathbf{u}) = 1/2^n \operatorname{Tr}(A_{\mathbf{u}}\rho)$, for all $\mathbf{u} \in V = \mathbb{Z}_2^{2n}$, where $A_{\mathbf{u}} = T_{\mathbf{u}} A_{\mathbf{0}} T_{\mathbf{u}}^\dagger$,

$$A_{\mathbf{0}} = \frac{1}{2^n} \sum_{\mathbf{a} \in V} T_{\mathbf{a}}. \tag{6}$$

This definition satisfies the minimal conditions required of a Wigner function [13], namely that (i) W is a quasi-probability distribution defined on a state space $V = \mathbb{Z}_2^{2n}$, (ii) W transforms covariantly under the Pauli group, $W_{T_{\mathbf{a}}\rho T_{\mathbf{a}}^\dagger}(\mathbf{u}) = W_\rho(\mathbf{u}+\mathbf{a})$, for all $\mathbf{u}, \mathbf{a} \in V$, and (iii) there is a suitable notion of marginals.

All previous works on the role of positive Wigner functions for QCSI—[10], [11], [9], [7]—are based on a particular family of Wigner functions for finite-dimensional state spaces introduced by Gibbons *et al.* [13]. This is, indirectly, also the case for the present Wigner function, and we therefore briefly describe its genealogy. Gibbons *et al.* introduced a family of Wigner functions for finite-dimensional state spaces based on the concepts of mutually unbiassed bases and lines in phase space. Among this family, for the special case of odd local dimension, Gross [14] identified a Wigner function which is the most sensible finite-dimensional analogue of the infinite-dimensional case [12]. This Wigner function was written in the form of Eqs. (5),(6) in [9], with a special phase convention γ. For local Hilbert space dimension 2, this special function γ does not exist, and in the present approach γ is left as a parameter to vary. The freedom of choosing the function γ replaces the freedom of choosing quantum nets in [13].

In addition to the above Properties (i) - (iii), the Wigner functions defined in Eqs. (5), (6) have two further relevant properties. First, for any pair ρ and σ of operators acting on the Hilbert space $\mathbb{C}^{2^n}$, it holds that

$$\operatorname{Tr}(\rho\sigma) = 2^n \sum_{\mathbf{u} \in V} W_\rho(\mathbf{u}) W_\sigma(\mathbf{u}). \tag{7}$$

Second, for any admissible function γ, we have the following relation between a quantum state ρ and its Wigner functions W_ρ,

$$\rho = \sum_{\mathbf{u} \in V} W_\rho(\mathbf{u}) A_{\mathbf{u}}.$$

Thus, the Wigner functions defined through Eqs. (5) and (6) all satisfy the constraint

(P2) Any n-qubit state ρ can be reconstructed from the corresponding Wigner function W_ρ.

There is an additional compatibility condition on the Wigner function W which has nothing to do with Wigner functions per se, but results from its intended use in the description of a QCSI scheme. Namely, the measurement of the observables in the set $\mathcal{O}$ must not introduce negativity into a formerly non-negative Wigner function,

(P3) If $W_\rho \geq 0$ then $W_{\frac{I\pm O}{2}\rho\frac{I\pm O}{2}} \geq 0$, for all $O \in \mathcal{O}$.

The motivation for enforcing property (P3) is that it gives rise to a notion of classicality based on efficient classical simulation of the free sector of QCSI, as will be discussed in Section IV. This feature is shared with the previously discussed cases of qudits [9], [6] and rebits [7]. Positivity of a Wigner function is generally associated with classicality; however, we can not refer to this viewpoint for motivation. The reason is that we do not require a counterpart of (P3) for the free unitaries. The unitaries in G *may* introduce negativity into the Wigner function, without affecting efficient classical simulability.

An immediate consequence of Property (P3) is that all free states $|\Psi\rangle \in \Omega$ are non-negatively represented by W. All free states $|\Psi\rangle \in \Omega$ can be created from the completely mixed state $I/2^n$, by measurement of observables in $\mathcal{O}$, and $W_{I/2^n} \geq 0$ for any γ. Then, with Property (P3), $W_{|\Psi\rangle\langle\Psi|} \geq 0$, for all $|\Psi\rangle \in \Omega$. By Eq. (2), free states remain non-negatively represented upon action of free unitary gates $g \in G$.

E. Two consistency conditions for the set $\mathcal{O}$

The set $\mathcal{O}$ of directly measurable observables is, in the present setting, always a set of Hermitian Pauli operators,

$$\mathcal{O} = \{\pm T_{\mathbf{a}}, \mathbf{a} \in V_{\mathcal{O}}\},$$

where $V_{\mathcal{O}}$ is a subset of $V = (\mathbb{Z}_2)^{2n}$.

In Section III C we described how to construct the set Ω of free states given the set $\mathcal{O}$ of directly measurable observables. But how is the set $\mathcal{O}$ itself constructed? To answer this question, we return to the function γ in Eq. (5) from which everything follows in the present setting. The function $\gamma : V \longrightarrow \mathbb{Z}_4$ specifies a function $\beta : V \times V \longrightarrow \mathbb{Z}_4$ defined via

$$T_{\mathbf{a}+\mathbf{b}} = i^{\beta(\mathbf{a},\mathbf{b})} T_{\mathbf{a}} T_{\mathbf{b}}. \tag{8}$$

The function β constrains the Pauli operators that can possibly be contained in the set $\mathcal{O}$. Namely, we have the following Lemma.

Lemma 1 *For any $\boldsymbol{a} \in V$, the measurement of an observable $\pm T_{\boldsymbol{a}}$ does not introduce negativity into the Wigner function if and only if*

$$\beta(\boldsymbol{a}, \boldsymbol{b}) = 0, \; \forall \boldsymbol{b} \in V \,|\, [\boldsymbol{a}, \boldsymbol{b}] = 0. \tag{9}$$

In Eq. (9), $[\cdot,\cdot]$ is the symplectic bilinear form defined by $[\mathbf{a}, \mathbf{b}] := \mathbf{a}_X \cdot \mathbf{b}_Z + \mathbf{a}_Z \cdot \mathbf{b}_X \mod 2$, for all $\mathbf{a}, \mathbf{b} \in V$.

Proof of Lemma 1. "Only if": Assume that the condition Eq. (9) does not hold, i.e., there exists a Pauli operator $T_{\mathbf{b}}$ such that $[\mathbf{a}, \mathbf{b}] = 0$ and $\beta(\mathbf{a}, \mathbf{b}) \neq 0 = 2$ (Hermiticity).

Further assume that the system is in the mixed state $(I - T_{\mathbf{b}})/2^n$, which has non-negative W, and that $T_{\mathbf{a}}$ is measured. W.l.o.g. assume that the outcome is -1. The resulting state is $\rho = (I - T_{\mathbf{a}} - T_{\mathbf{b}} + T_{\mathbf{a}} T_{\mathbf{b}})/2^n = (I - T_{\mathbf{a}} - T_{\mathbf{b}} - T_{\mathbf{a}+\mathbf{b}})/2^n$. Thus, $W_\rho(\mathbf{0}) = -2/4^n < 0$. Thus, if $\beta(\mathbf{a}, \mathbf{b}) \neq 0$ for some $\mathbf{b} \in V$, the measurement of $T_{\mathbf{a}}$ can introduce negativity into Wigner functions, hence $\pm T_{\mathbf{a}} \notin \mathcal{O}$. Negation of this statement proofs the result.

"If": We assume that the Wigner function W_ρ of the state ρ before the measurement is non-negative,

$$W_\rho(\mathbf{u}) \geq 0, \; \forall \mathbf{u} \in V,$$

and that the measured observable $T_{\mathbf{a}}$ is such that $\beta(\mathbf{a}, \mathbf{b}) = 0$, for all $\mathbf{b} \in V$. The state ρ' after the measurement of the observable $T_{\mathbf{a}}$ with outcome $s \in \{0, 1\}$ is $\rho' \sim \frac{I+(-1)^s T_{\mathbf{a}}}{2} \rho \frac{I+(-1)^s T_{\mathbf{a}}^\dagger}{2}$, and the value of the corresponding Wigner function at the phase space point $\mathbf{u} \in V$ is

$$p_{\mathbf{a}}(s) W_{\rho'}(\mathbf{u}) = \frac{1}{2^n} \mathrm{Tr}\left(\frac{I + (-1)^s T_{\mathbf{a}}^\dagger}{2} A_{\mathbf{u}} \frac{I + (-1)^s T_{\mathbf{a}}}{2} \rho \right). \tag{10}$$

Therein, $p_{\mathbf{a}}(s)$ is the probability of obtaining the outcome s in the measurement of $T_{\mathbf{a}}$. Now,

$$\begin{aligned}
&\frac{I + (-1)^s T_{\mathbf{a}}^\dagger}{2} A_{\mathbf{u}} \frac{I + (-1)^s T_{\mathbf{a}}}{2} = \\
&= \frac{I + (-1)^s T_{\mathbf{a}}^\dagger}{2} \left(\frac{1}{2^n} \sum_{\mathbf{b} \in V} (-1)^{[\mathbf{u},\mathbf{b}]} T_{\mathbf{b}} \right) \frac{I + (-1)^s T_{\mathbf{a}}}{2} \\
&= \frac{I + (-1)^s T_{\mathbf{a}}}{2} \left(\frac{1}{2^n} \sum_{\mathbf{b} \in V \,|\, [\mathbf{b},\mathbf{a}]=0} (-1)^{[\mathbf{u},\mathbf{b}]} T_{\mathbf{b}} \right) \\
&= \frac{1}{2^{n+1}} \sum_{\mathbf{b} \in V \,|\, [\mathbf{b},\mathbf{a}]=0} (-1)^{[\mathbf{u},\mathbf{b}]} \big(T_{\mathbf{b}} + (-1)^s T_{\mathbf{a}} T_{\mathbf{b}} \big) \\
&= \frac{1}{2^{n+1}} \sum_{\mathbf{b} \in V \,|\, [\mathbf{b},\mathbf{a}]=0} (-1)^{[\mathbf{u},\mathbf{b}]} \big(T_{\mathbf{b}} + (-1)^s T_{\mathbf{a}+\mathbf{b}} \big) \\
&= \frac{1}{2^{n+1}} \sum_{\mathbf{b} \in V \,|\, [\mathbf{b},\mathbf{a}]=0} (-1)^{[\mathbf{u},\mathbf{b}]} \left(1 + (-1)^{s + [\mathbf{a},\mathbf{u}]} \right) T_{\mathbf{b}} \\
&= \frac{\delta_{s,[\mathbf{a},\mathbf{u}]}}{2^n} \sum_{\mathbf{b} \in V \,|\, [\mathbf{b},\mathbf{a}]=0} (-1)^{[\mathbf{u},\mathbf{b}]} T_{\mathbf{b}} \\
&= \frac{\delta_{s,[\mathbf{a},\mathbf{u}]}}{2} (A_{\mathbf{u}} + A_{\mathbf{u}+\mathbf{a}}).
\end{aligned}$$

Above, we have used the assumption that $\beta(\mathbf{a}, \mathbf{b}) = 0$ for all $\mathbf{b} \in V$ when transitioning from line 4 to line 5. Applying the result to Eq. (10) we find that

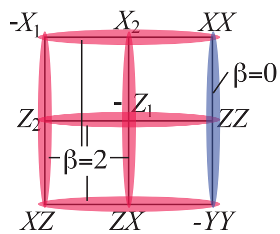
$$p_{\mathbf{a}}(s) W_{\rho'}(\mathbf{u}) = \frac{\delta_{s,[\mathbf{a},\mathbf{u}]}}{2} \big(W_\rho(\mathbf{u}) + W_\rho(\mathbf{u} + \mathbf{a}) \big). \tag{11}$$

By assumption, the r.h.s. is always non-negative. For the outcome s to possibly occur, it is required that $p_{\mathbf{a}}(s) > 0$. Hence, $W_{\rho'}(\mathbf{u}) \geq 0$, for all $\mathbf{u} \in V$. □

Example: To illustrate the usefulness of Lemma 1, consider the following choice for γ. For brevity, we restrict to two rebits. W^γ is specified by

$$\begin{aligned}
A_{\mathbf{0}}^\gamma \;=\; & \tfrac{1}{4}\big(I - Z_1 + Z_2 + Z_1 Z_2 - X_1 + X_2 + X_1 X_2 + \\
& + X_1 Z_2 + Z_1 X_2 - Y_1 Y_2 \big).
\end{aligned}$$

Arranging all observables in A_0 apart from the identity in Mermin's square,



it is evident that every observable $T_{\mathbf{a}}$ is part of at least one commuting triple with $\beta \neq 0$. Hence, apart from the identity, no observable is in $\mathcal{O}$, i.e., $\mathcal{O} = \{I\}$. The corresponding QCSI scheme is thus the exact opposite of tomographically complete: Nothing can be measured at all! We find that not for every function γ the Wigner function W^γ can be paired with a matching QCSI scheme.

For further illustration of Lemma 1, we have the following implication.

Lemma 2 *Consider a Wigner function as defined in Eqs. (6), (5), for $n \geq 2$ qubits. Then, there always exists a Pauli observable whose measurement does not preserve positivity.*

Thus, no QCSI scheme in which all Pauli observables are directly measurable can satisfy the property (P3). The original QCSI scheme [8] on qubits is one of those schemes, and it is therefore out of scope of the present analysis.

Proof of Lemma 2. Consider Mermin's square as displayed in Fig. 1. Irrespective of the sign conventions of the Pauli observables contained in it, there is always at least one context with $\beta = 2$. Thus, by Lemma 1, for $n \geq 2$ the measurement of at least three Pauli observables introduces negativity into the Wigner function. □

Returning to the question of which constraints exist for possible sets $\mathcal{O}$ of directly measurable observables, there is one more. It stems from the requirement of tomographic completeness of QCSI.

Definition 1 *$M = \{\pm T_{\boldsymbol{a}} | \, \boldsymbol{a} \in V_M\}$ is the set of Pauli observables whose value can be inferred from a* single copy *of the given quantum state, by measurement of other Pauli observables and classical post-processing.*

The set M is typically larger than the set $\mathcal{O}$ of observables which can be directly measured. This was illustrated by an example in Section III B, namely $\mathcal{O} = \{I, X_1, Z_2\}$, $M = \mathcal{O} \cup \{X_1 Z_2\}$. We now provide a general characterization of the set M generated by the set $\mathcal{O}$.

Lemma 3 *For any γ, the set V_M has the properties that (i) $V_{\mathcal{O}} \subseteq V_M$, and (ii) for any $\boldsymbol{a} \in V_{\mathcal{O}}$, $\boldsymbol{b} \in V$ with $[\boldsymbol{a}, \boldsymbol{b}] = 0$, it holds that $\boldsymbol{a} + \boldsymbol{b} \in V_M$ if and only if $\boldsymbol{b} \in V_M$.*

Proof of Lemma 3. Property (i) merely states that what can be directly measured can have its value inferred. Regarding (ii), the observable $T_{\mathbf{a}+\mathbf{b}}$ has its value inferred as follows. First, $T_{\mathbf{a}} \in \mathcal{O}$ is measured directly. Then, the procedure for inferring the value of $T_{\mathbf{b}}$ is applied. Since $T_{\mathbf{a}}$ commutes with $T_{\mathbf{b}}$, the former measurement doesn't interfere with the latter, and $\mu(T_{\mathbf{a}} T_{\mathbf{b}}) = \mu(T_{\mathbf{a}})\mu(T_{\mathbf{b}})$. Finally, with Eq. (9), $\mu(T_{\mathbf{a}+\mathbf{b}}) = \mu(T_{\mathbf{a}})\mu(T_{\mathbf{b}})$. Thus, if $\mathbf{b} \in V_M$ then $\mathbf{a} + \mathbf{b} \in V_M$. The reverse direction holds by symmetry in $\mathbf{b} \longleftrightarrow \mathbf{a} + \mathbf{b}$. □

Example. Assume that $X_1, Z_2, Y_1 Y_2 \in \mathcal{O}$. The outcome of the observable $Z_1 X_2$ can then be inferred by measurement, i.e. $Z_1 X_2 \in M$. The procedure for the measurement of the observable $Z_1 X_2$, given the above set $\mathcal{O}$ of directly measurable observables, is the following. First, the observable $Y_1 Y_2$ is measured, and second the commuting observables X_1 and Z_2 are measured. The measurement outcome $\mu(Z_1 X_2) \in \{\pm 1\}$ then is

$$\mu(Z_1 X_2) = \mu(Y_1 Y_2)\mu(X_1)\mu(Z_2).$$

The key point of this example is that not all pairs among the measured Pauli observables X_1, Z_2 and $Y_1 Y_2$ commute; yet in the above expression for $\mu(Z_1 X_2)$ we treated them as if they did. How is that possible?

Since $Y_1 Y_2$ does not commute with X_1 and Z_2, the measurements of X_1 and Z_2 after the measurement of $Y_1 Y_2$—if taken separately—do not reveal any information about the initial state. Individually, their outcomes are completely random, whatever the state prior to the $Y_1 Y_2$-measurement is. However, X_1 and Z_2 mutually commute, and hence the separate measurement of X_1 and Z_2 implies a valid measurement outcome for the correlated observable $X_1 Z_2$, namely $\mu(X_1 Z_2) = \mu(X_1)\mu(Z_2)$. Furthermore, since $X_1 Z_2$ does commute with $Y_1 Y_2$, $\mu(X_1)\mu(Z_2)$ represents the outcome of a $X_1 Z_2$-measurement on the initial state, and $\mu(Z_1 X_2) = \mu(Y_1 Y_2)\mu(X_1 Z_2) = \mu(Y_1 Y_2)\mu(X_1)\mu(Z_2)$, as claimed.

Let us now verify that $Z_1 X_2 \in M$ follows from the properties established in Lemma 3. First, with Property (i) of Lemma 3, $X_1 \in \mathcal{O}$ implies $X_1 \in M$. Then, using Property (ii) with $X_1 \in M$, $Z_2 \in \mathcal{O}$, it follows that $X_1 Z_2 \in M$. Finally, again with Property (ii), since $Y_1 Y_2 \in \mathcal{O}$ and $X_1 Z_2 \in M$, it follows that $Z_1 X_2 \in M$.

We note that the above procedure of inferring measurement outcomes by the physical measurement of non-commuting observables is reminiscent of the syndrome measurement in subsystem codes [33], [34], with the Bacon-Shor code [35], [36] and topological subsystem codes [37] as prominent examples.

Back to the general scenario, an observable $T_{\mathbf{a}}$ can have its value inferred, i.e., $T_{\mathbf{a}} \in M$, if there exists a resolution

$$\mathbf{a} = \mathbf{a}_1 + (\mathbf{a}_2 + (\mathbf{a}_3 + ...(\mathbf{a}_{N-1} + \mathbf{a}_N)...)) \,, \tag{12}$$

where all $\mathbf{a}_i \in V_{\mathcal{O}}$, and

$$\left[\mathbf{a}_i, \sum_{j=i+1}^{N} \mathbf{a}_j \right] = 0, \; \forall i = 1,..,N-1. \tag{13}$$

The resolution Eq. (12) of $\mathbf{a}$ represents a measurement sequence for inferring the value of $T_\mathbf{a}$, starting with the measurement of $\mathbf{a}_1$ and ending with the measurement of $\mathbf{a}_N$. The inferred value is $\lambda(T_\mathbf{a}) = \prod_{i=1}^N \lambda(T_{\mathbf{a}_i})$.

In addition, we observe that Lemma 3 can be turned into a procedure for finding V_M given $V_\mathcal{O}$ as a generating set. The procedure is the following.

1. Set $V_M := V_\mathcal{O}$. Then repeat the subsequent steps 2 and 3 until V_M saturates:
2. Construct the set
$$\Delta V_M := \{\mathbf{a}+\mathbf{b} |\, \mathbf{a} \in V_\mathcal{O}, \mathbf{b} \in V_M, [\mathbf{a},\mathbf{b}] = 0\}. \tag{14}$$
3. Update $V_M \longrightarrow V_M \cup \Delta V_M$.

Since $V_M \subseteq V$ is finite, this algorithm always terminates.

The condition (P1) of tomographic completeness reads
$$V_M = V. \tag{15}$$

Thus far, we have identified two constraints on the set $\mathcal{O}$ of directly measurable observables, Eq. (9) and Eq. (15). Eq. (9) is the necessary and sufficient condition for the measurement of the observable $T_\mathbf{a}$ to not introduce negativity into a non-negative Wigner function. Eq. (15) is the necessary and sufficient condition for tomographic completeness. Eqs. (9), (15) thus represent the requirements (P1) and (P3).

In the following, we will be mostly interested in the non-extendable solutions $\mathcal{O}$ of Eqs. (9) and (15).

Definition 2 *A set $\mathcal{O}$ of directly measurable observables is* consistent *if it satisfies Eqs. (9) and (15). A consistent set $\mathcal{O}$ is* non-extendable *if no consistent set $\mathcal{O}'$ exists such that $\mathcal{O} \subsetneq \mathcal{O}'$.*

Lemma 4 *For any given function γ, if Eqs. (9), (15) have a solution, then the non-extendable solution is unique. It is*
$$\mathcal{O}_{max} = \{\pm T_{\boldsymbol{a}},\, \boldsymbol{a} \in V |\, \beta(\boldsymbol{a},\boldsymbol{u}) = 0,\, \forall \boldsymbol{u} \in V_{\boldsymbol{a}}\}, \tag{16}$$
with $V_{\boldsymbol{a}} := \{\boldsymbol{u} \in V | [\boldsymbol{a},\boldsymbol{u}] = 0\}$.

Remark: At least for some γ, already the consistent set $\mathcal{O}$ is unique, without requiring non-extendability to begin with. Section VI B presents an example.

Proof of Lemma 4. First, we show that for any consistent $\mathcal{O}'$ it holds that $\mathcal{O}' \subseteq \mathcal{O}_\text{max}$.

The proof is indirect. Assume there exists a consistent set $\mathcal{O}'$ such that $\mathcal{O}' \not\subseteq \mathcal{O}_\text{max}$. Then there exists an $\mathbf{a} \in V$ such that $T_\mathbf{a} \in \mathcal{O}'$ and $T_\mathbf{a} \notin \mathcal{O}_\text{max}$. Then, by the definition Eq. (16) of $\mathcal{O}_\text{max}$, there exists a $\mathbf{u} \in V$ such that $[\mathbf{a},\mathbf{u}] = 0$ and $\beta(\mathbf{a},\mathbf{u}) \neq 0$. Then, with Lemma 1, the measurement of $T_\mathbf{a}$ has the power to introduce negativity into previously non-negative Wigner functions. Contradiction. Thus, if $\mathcal{O}_\text{max}$ is consistent, then it is the unique non-extendable consistent set.

Second, we show that if there is any consistent set $\mathcal{O}'$, then $\mathcal{O}_\text{max}$ is consistent.

The set $\mathcal{O}_\text{max}$ satisfies Property (P3) by construction, cf. Lemma 1. The only way for it to fail consistency is by failing Property (P1) of tomographic completeness, cf. Eq. (15). Assume it does fail Eq. (15), i.e., $V_M(\mathcal{O}_\text{max}) \subsetneq V$. Then, since every consistent solution $\mathcal{O}'$ has the property $\mathcal{O}' \subseteq \mathcal{O}_\text{max}$, with Eq. (14) it follows that $V_M(\mathcal{O}') \subseteq V_M(\mathcal{O}_\text{max})$. Hence, $V_M(\mathcal{O}') \subsetneq V$, obstructing consistency. Thus, if $\mathcal{O}_\text{max}$ is not consistent, then no consistent set $\mathcal{O}'$ exists.

Chaining the above two statements together, if a consistent set $\mathcal{O}'$ exists, then $\mathcal{O}_\text{max}$ is consistent, and then it is the unique non-extendable consistent set. □

Finally, we introduce a generalization of the set M of Pauli observables whose value can be inferred. Namely, we denote by C, $C \subset M$, a set of observables whose value can be inferred *jointly* in QCSI. For short, we call such a set C a "set of jointly measurable observables".

Definition 3 *A set C, $C \subset M$, of commuting Pauli observables is jointly measurable if the outcomes for all observables in C can be simultaneously inferred from measurement of observables in $\mathcal{O}$ on a* single copy *of the given quantum state ρ.*

The sets C of simultaneously measurable observables will become important in Section V C, where we discuss the relation between contextuality and negativity of Wigner functions.

Some examples for possible sets C are (i) $C = \{O\}$, for any $O \in M$, (ii) any commuting subset of $\mathcal{O}$, and (iii) $C = \{A, B, AB\}$, for $A \in M$, $B \in \mathcal{O}$ and $[A, B] = 0$.

We have the following characterization of the sets C of simultaneously measurable observables.

Lemma 5 *Consider a set C of simultaneously measurable observables, and $T_{\boldsymbol{a}}, T_{\boldsymbol{b}} \in C$. Then, $T_{\boldsymbol{a}+\boldsymbol{b}} = T_{\boldsymbol{a}} T_{\boldsymbol{b}}$.*

We postpone the proof of Lemma 5 until Section V B.

F. Constructing QCSI schemes from γ

As stated above, for any QCSI scheme on qubits we require that the available measurements are tomographically complete, cf. (P1), and that the allowed Pauli measurements do not introduce negativity into the Wigner function, cf. (P3). These two requirements are potentially at tension. The first requirement is easier to satisfy for large sets $\mathcal{O}$ of directly measurable observables, whereas the second requirement is easier to satisfy for a small such set. It is not a priori clear that a set $\mathcal{O}$ exists which satisfies both requirements.

Once γ is given, the Wigner function and the free sector of the corresponding non-extendable QCSI scheme are fully specified. They are obtained through the following steps:

1. Construct the Wigner function W via its definition Eqs. (5), (6).
2. Compute the function β defined in Eq. (8) from the function γ. Construct the maximal set $\mathcal{O}_{\text{max}}$ of directly measurable observables via Eq. (16).
3. Using the algorithm of Eq. (14), construct the set M generated by $\mathcal{O}_{\text{max}}$. Check that the resulting measurements lead to tomographically complete QCSI, i.e. that Eq. (15) is satisfied.
4. Construct the group G of free unitary gates via Eq. (3), and set Ω of free states via Eq. (4).

The test in Step 3 fails for some functions γ. For those Wigner functions W^γ there simply is no corresponding tomographically complete QCSI scheme. The magic states do not follow from the present analysis, and need to be constructed case by case.

To summarize, in this section we have stated minimal requirements for any QCSI scheme on qubits and its corresponding Wigner function. We have shown that once the function γ is provided, the corresponding non-extendable QCSI scheme—if it exists—is fully specified.

IV. EFFICIENT CLASSICAL SIMULATION OF QCSI FOR NON-NEGATIVE STATES

The requirement (P3) of positivity preservation of the Wigner function has a bearing on classical simulation of QCSI, which we describe in this section. Below we present an efficient classical simulation method based on sampling, which leads to the following result.

Theorem 1 *Consider a QCSI with input state ρ_{in}, fed into a sequence of unitary gates $g \in G$ and measurements of observables in $\mathcal{O}$. If (i) the Wigner function $W_{\rho_{in}}$ of ρ_{in} can be efficiently sampled from, and (ii) the phase convention $\gamma(\boldsymbol{a})$ can be efficiently evaluated for all $\boldsymbol{a} \in V_{\mathcal{O}}$, then the distribution of measurement outcomes can be efficiently sampled from.*

A. Reformulation of the simulation problem

For the purpose of classical simulation, we make an alternation to the present QCSI scheme, which, however, does not affect its computational power. Namely, we absorb the unitary gates in the measurements, such that only state preparations and measurements remain as free operations. Here we take the viewpoint that all there is to simulate about a quantum computation is the joint outcome distribution of measurements performed in course of the computation. If the unitaries can be removed without altering the outcome distribution, then

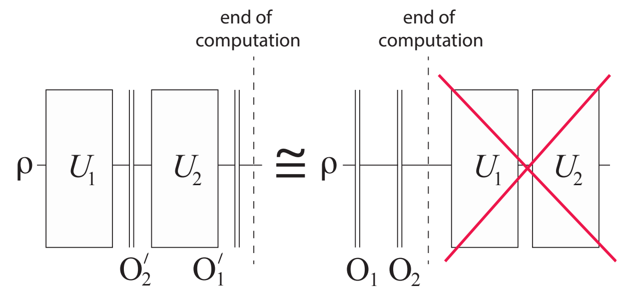


FIG. 2: The measurements of observables O_i' are propagated backwards in time to act on the initial state, by conjugation under the interspersed unitaries. Since only the measurement statistics is of interest, the trailing unitaries may be removed from the resulting circuit.

there is certainly no loss in removing them. But there is a gain: as is made explicit in Section IV C, the simulation algorithm of Section IV B can handle free unitaries $g \in G$ that introduce negativity into the Wigner function of the processed quantum state.

The general procedure is outlined in Fig. 2. Consider a QCSI circuit which is an alternation of unitary gates $g_i(\mathbf{s}_{\prec i})$ and projective measurements represented by projectors $P_i'(\mathbf{s}_{\prec i}, s_i)$,

$$\mathcal{C} = \prod_{i=1}^{t_{\max}} P_i'(\mathbf{s}_{\prec i}, s_i) g_i(\mathbf{s}_{\prec i}), \tag{17}$$

where the factors in the product are ordered from right to left. Therein, $\mathbf{s}$ is the binary vector of all measurement outcomes, and $\mathbf{s}_{\prec i}$ is $\mathbf{s}$ restricted to the measurement outcomes obtained prior to the gate g_i. We thus allow unitary gates and measurements to depend on previously obtained measurement outcomes. Such conditioning is essential for the working of QCSI.

Now denote by $G_i(\mathbf{s}_{\prec i})$ the unitaries accumulated up to step i,

$$G_i(\mathbf{s}_{\prec i}) = \prod_{j=1}^{i} g_i(\mathbf{s}_{\prec i}). \tag{18}$$

Therein, the ordering of operations is the same as in Eq. (17). The circuit $\mathcal{C}$ of Eq. (17) may then be rewritten as

$$\begin{aligned} \mathcal{C} &= \prod_{i=1}^{t_{\max}} P_i'(\mathbf{s}_{\prec i}, s_i) g_i(\mathbf{s}_{\prec i}) \\ &= G_{t_{\max}}(\mathbf{s}) \prod_{i=1}^{t_{\max}} G_i(\mathbf{s}_{\prec i})^\dagger P_i'(\mathbf{s}_{\prec i}, s_i) G_i(\mathbf{s}_{\prec i}) \\ &\cong \prod_{i=1}^{t_{\max}} G_i(\mathbf{s}_{\prec i})^\dagger P_i'(\mathbf{s}_{\prec i}, s_i) G_i(\mathbf{s}_{\prec i}). \end{aligned}$$

Thus, if the measured observables in the original sequence of operations were $O_i'(\mathbf{s}_{\prec i})$, the corresponding observables in the equivalent sequence are

$$O_i(\mathbf{s}_{\prec i}) = G_i(\mathbf{s}_{\prec i})^\dagger O_i'(\mathbf{s}_{\prec i}) G_i(\mathbf{s}_{\prec i}). \tag{19}$$

By Eq. (1), if $O'_i(\mathbf{s}_{\prec i}) \in \mathcal{O}$, then $O_i(\mathbf{s}_{\prec i}) \in \mathcal{O}$. Therefore, a QCSI scheme with set $\mathcal{O}$ of measurable observables and group G of unitary gates is equivalent to a QCSI scheme with set $\mathcal{O}$ of measurable observables and no unitaries at all.

B. Simulation algorithm

The classical simulation algorithm for the setting of Theorem 1 is the following.

Algorithm 1

1. Draw a sample $\mathbf{v} \in V$ from $W_{\rho_{\text{in}}}$, and set $\mathbf{v}_1 := \mathbf{v}$.
2. For all the measurements of observables $T_{\mathbf{a}_i} \in \mathcal{O}$ comprising the circuit, starting with the first,
   (a) Output the result $s_i = [\mathbf{a}_i, \mathbf{v}_i]$ for the measurement of the observable $T_{\mathbf{a}_i}$,
   (b) Flip a fair coin, and update the sample
   $$\mathbf{v}_i \longrightarrow \mathbf{v}_{i+1} = \begin{cases} \mathbf{v}_i, & \text{if "heads"} \\ \mathbf{v}_i + \mathbf{a}_i, & \text{if "tails'} \end{cases},$$
   until the measurement sequence is complete.
3. Repeat until sufficient statistics is gathered.

Any sample $\mathbf{u} \in V$ from a non-negative Wigner function has a definite value assignment for all observables in the Pauli group. Namely, for any $\mathbf{a} \in V$, the measurement outcome for the Pauli observable $T_\mathbf{a}$ is

$$\lambda_\mathbf{u}(\mathbf{a}) = (-1)^{[\mathbf{a},\mathbf{u}]}. \tag{20}$$

The value assignment Eq. (20) is a direct consequence of the update rule Eq. (11). Namely, in the l.h.s. of Eq. (11) assume that the probability $p_\mathbf{a}(s)$ for obtaining the outcome $s \in \{0,1\}$ in a measurement of a Pauli observable $T_\mathbf{a}$ on a state ρ is non-zero. Then, the r.h.s. of Eq. (11) implies that $s = [\mathbf{a},\mathbf{u}]$, or, equivalently, $\lambda_\mathbf{u} = (-1)^{[\mathbf{a},\mathbf{u}]}$.

For illustration of the state update rule in the above classical simulation algorithm, consider two measurement sequences for the state of one qubit, namely (i) Repeated measurements of the Pauli observable $Z = \pm T_{\mathbf{a}(Z)}$, and (ii) Alternating measurements of the Pauli observables Z and $X = \pm T_{\mathbf{a}(X)}$. Assume that the sample from the Wigner function of the initial state is $\mathbf{u} \in V$. Regarding (i), according to the classical simulation algorithm, the ontic states after one or a larger number of measurements are $\mathbf{u}$ or $\mathbf{u} + \mathbf{a}(Z)$. Either way, the reported measurement outcome is $(-1)^{[\mathbf{u},\mathbf{a}(Z)]}$, since $[\mathbf{a}(Z),\mathbf{a}(Z)] = 0$. For any sample $\mathbf{u}$ from the input distribution, the sequence of measurement outcomes is thus constant, as required. Regarding (ii), since $[\mathbf{a}(Z),\mathbf{a}(X)] = 1$, from the second measurement onwards the outcomes produced by the classical simulation are completely random and uncorrelated, as required.

Proof of Theorem 1. The theorem follows from the efficiency and the correctness of the above classical simulation algorithm.

Efficiency: Both the pre-processing of removing the unitaries from the circuit and the classical simulation algorithm itself need to be considered.

(1) Pre-processing. We need to track the evolution of Pauli observables $T_\mathbf{a}$ under conjugation by gates $g \in G$, as described in Eq. (19). This can be done efficiently within the stabilizer formalism [23]. However, the stabilizer formalism uses its own phase convention $\tilde{\gamma}$ for the Pauli operators, $\tilde{T}_\mathbf{a} := i^{\tilde{\gamma}(\mathbf{a})} Z(\mathbf{a}_Z)X(\mathbf{a}_X)$, such that $\tilde{\gamma}$ can be efficiently evaluated. Suppose, $g^\dagger \tilde{T}_\mathbf{a} g = i^{\tilde{\phi}_g(\mathbf{a})}\tilde{T}_{g(\mathbf{a})}$. Then, $g^\dagger T_\mathbf{a} g = i^{\phi_g(\mathbf{a})} T_{g(\mathbf{a})}$, with

$$\phi_g(\mathbf{a}) = \tilde{\phi}_g(\mathbf{a}) + (\tilde{\gamma}(g\mathbf{a}) - \tilde{\gamma}(\mathbf{a})) - (\gamma(g\mathbf{a}) - \gamma(\mathbf{a}))\,.$$

By assumption, γ can be efficiently evaluated, and hence can ϕ_g, for any $g \in G$.

(2) Classical simulation algorithm. The efficiency of the above classical simulation algorithm is evident.

Correctness: Assume that the classical simulation algorithm samples correctly from the Wigner function of the state ρ_t after the t-th measurement in the sequence. We now show that under this assumption (i) The above classical simulation algorithm produces the correct probability distribution for the $(t+1)$-th measurement, and (ii) correctly samples from the Wigner function of the conditional state $\rho_{t+1}(s_{t+1})$ after the $(t+1)$-th measurement.

(i) According to the value assignment Eq. (20) of the classical simulation algorithm, the probability $p_\mathbf{a}(s)$ for obtaining the outcome $s \in \{0,1\}$ in the measurement of the observable $T_\mathbf{a}$ on the state ρ_t is

$$p_\mathbf{a}(s) = \sum_{\mathbf{u}\in V} \delta_{[\mathbf{a},\mathbf{u}],s} W_{\rho_t}(\mathbf{u}).$$

As is easily verified by direct calculation,

$$W_{\frac{I+(-1)^s T_\mathbf{a}}{2}}(\mathbf{u}) = \frac{1}{2^n}\delta_{[\mathbf{a},\mathbf{u}],s}. \tag{21}$$

Combining the last two equations, and using the property Eq. (7), we find that

$$\begin{aligned} p_\mathbf{a}(s) &= 2^n \textstyle\sum_{\mathbf{u}\in V} W_{\frac{I+(-1)^s T_\mathbf{a}}{2}}(\mathbf{u}) W_{\rho_t}(\mathbf{u}) \\ &= \operatorname{Tr}\left(\frac{I+(-1)^s T_\mathbf{a}}{2}\rho_t\right), \end{aligned}$$

which is the quantum-mechanical expression.

(ii) Consider the Wigner function W_{ρ_t} for the state ρ_t after step t in the expansion $W_{\rho_t} = \sum_{\mathbf{u}\in V} W_{\rho_t}(\mathbf{u})\delta_\mathbf{u}$. At time $t+1$, the observable $T_\mathbf{a}$ is measured, with outcome $s \in \{0,1\}$. Then, from the value assignment Eq. (20), only the phase space points $\mathbf{u} \in V$ with $s = [\mathbf{u}, \mathbf{a}]$ contribute to conditional density matrix $\rho_{t+1}(s)$. Furthermore, per Step 2b of the classical simulation algorithm, the update for δ-distributions over phase space is $\delta_\mathbf{u} \mapsto (\delta_\mathbf{u} + \delta_{\mathbf{u}+\mathbf{a}})/2$. Therefore, the Wigner function for the (normalized) conditional state $\rho_{t+1}(s)$, according to the classical simulation algorithm, is

$$p_\mathbf{a}(s) W_{\rho_{t+1}(s)} = \sum_{\mathbf{u}\in V} \delta_{[\mathbf{a},\mathbf{u}],s} W_{\rho_t}(\mathbf{u}) \frac{\delta_\mathbf{u} + \delta_{\mathbf{u}+\mathbf{a}}}{2}.$$

Hence, $p_\mathbf{a}(s) W_{\rho_{t+1}(s)}(\mathbf{v}) = \delta_{[\mathbf{a},\mathbf{v}],s} \frac{W_{\rho_t}(\mathbf{v}) + W_{\rho_t}(\mathbf{v}+\mathbf{a})}{2}$, which is the quantum-mechanical expression Eq. (11).

By assumption of Theorem 1, the Wigner function of the initial state $\rho_\text{in} = \rho_0$ is correctly sampled from. Thus, with the above statements (i) and (ii), it follows by induction that all sequences of measurement outcomes occur with the correct probabilities. $\Box$

C. Discussion

A notable property of the above simulation method is that, for any Wigner function employed therein, covariance and preservation of positivity under the group G of free gates are not required. This is a consequence of the reformulation of QCSI in Section IV A, where the free unitary gates were eliminated. It is in sharp contrast to the previously considered cases of qudits [9],[6] and rebits [7], where covariance and preservation of positivity under G were critical for the classical simulation by sampling. These points, and the roles remaining for covariance and positivity preservation in the present simulation method are discussed below.

1. *Covariance*

As an example, consider a quantum circuit for a single qubit consisting of a Hadamard gate followed by a measurement of the Pauli observable Z. Given is a source that samples from the non-negative Wigner function of the initial state ρ_in, and the task is to sample from the output distribution of the measurement.

A classical simulation method based on Wigner function covariance would, in the first step, convert the source that samples from W_{ρ_in} into a source that samples from $W_{H\rho_\text{in}H^\dagger}$, using covariance. In the second step, it would, for each sample $\mathbf{u}$ drawn from $W_{H\rho_\text{in}H^\dagger}$, output the value $(-1)^{[\mathbf{a}(Z),\mathbf{u}]}$, with $\mathbf{a}(Z)$ such that $T_{\mathbf{a}(Z)} = Z$; cf. Eq. (20). But there is a problem:

Lemma 6 *For any number n of qubits, no Wigner function of the type defined in Eqs. (5), (6) is covariant under a Hadamard gate on a single qubit.*

Step 1 of the above procedure cannot be performed!

Proof of Lemma 6. We only discuss $n = 1$, the generalization to other n is straightforward. Consider the phase point operator $A_0 = (I + i^{\gamma_x} X + i^{\gamma_y} Y + i^{\gamma_z} Z)/2$. For W to be covariant under H, we require that $H^\dagger A_0 H = A_\mathbf{u}$, for some $\mathbf{u}$. Now consider the sum of signs $\eta = \gamma_x + \gamma_y + \gamma_z \mod 4$, and how it transforms under H. Since $H^\dagger X H = Z$, $H^\dagger Y H = -Y$, and $H^\dagger Z H = X$, it follows that $\eta \longrightarrow \eta' = \eta + 2 \mod 4$. However, under the transformation $A_0 \longrightarrow A_\mathbf{u} = T_\mathbf{u}^\dagger A_0 T_\mathbf{u}$, the signs of an even number of $\{X, Y, Z\}$ are flipped, hence η remains unchanged mod 4. Thus $H^\dagger A_0 H \neq A_\mathbf{u}$ for any $\mathbf{u}$, for any γ. $\Box$

On the other hand, the simulation method described in the previous section has no problem with the above 1-qubit circuit. This example illustrates the fact that the simulation method of Section IV A is more widely applicable than simulation methods employing Wigner function covariance.

Nonetheless, the covariance group H of the Wigner function retains some relevance for classical simulation. Namely, it is the group of transformations that leave the simulation method of Section IV A invariant. Specifically, consider the joint transformation

$$\begin{aligned} \rho_\text{in} &\longrightarrow h^\dagger \rho_\text{in} h, \\ T_\mathbf{a} &\longrightarrow h^\dagger T_\mathbf{a} h,\ \forall T_\mathbf{a} \in \mathcal{O}_\text{max}, \end{aligned}$$

where $h \in H$. By unitarity, the transformation h does not change the outcome distribution. Also, it does not affect the simulation. First, the original sampling source is efficiently transformed into another sampling source; and second, the measurements remain in the efficiently simulable set $\mathcal{O}_\text{max}$. The latter holds because, by definition, H leaves β invariant. Hence, $\mathcal{O}_\text{max}$, which is defined through Eq. (16), remains unchanged. Thus, $H \subset G$ always holds.

In the previously discussed cases of qudits [9] and rebits [7], the covariance group H was synonymous with the group G of free gates. However, in general, $G = H$ doesn't have to hold, and in Section VI we present an example where it indeed doesn't hold.

As a consequence, the covariance group H of the Wigner function is becoming less of a dynamical concept for QCSI. It remains the invariance group of the classical simulation algorithm based on sampling from the Wigner function.

2. *Preservation of positivity*

As an example, consider a quantum circuit for two qubits consisting of a Hadamard gate H_1 on the first qubit followed by a measurement of the Pauli observable Z_1. Assume the initial state ρ_in is the completely mixed state, for which each Wigner function of the type Eq. (6) is positive and can be efficiently sampled from. The task is to sample from the output distribution of the measurement.

Again, a classical simulation method based on the preservation of Wigner function positivity under free unitaries again runs into a problem:

Lemma 7 *For $n \geq 2$, for no Wigner function of type Eq. (6) positivity is preserved for all states under a Hadamard gate on a single qubit.*

Proof of Lemma 7. Consider a real stabilizer state ρ of two qubits, $\rho = (I + T_\mathbf{a} + T_\mathbf{b} + T_\mathbf{a} T_\mathbf{b})/4$, and w.l.o.g the Hadamard gate H_1 on the first qubit. As a consequence of Lemma 1, $W_\rho \geq 0$ if and only if $\beta(\mathbf{a}, \mathbf{b}) = 0$. Likewise, for the transformed state, $W_{H_1 \rho H_1^\dagger} \geq 0$ if and only if $\beta(H_1 \mathbf{a}, H_1 \mathbf{b}) = 0$. Now consider Mermin's square. There are six contexts, i.e., sets of commuting Pauli observables such that within each set the observables multiply to the identity times ± 1. Whatever the phase convention γ for the nine Pauli operators in the square, there is always an odd number of contexts for which $\beta \mod 4 = 2$. Namely, for the standard phase convention, there is one such context. If the sign of any of the Pauli observables is flipped, then $\beta \longrightarrow \beta + 2 \mod 4$ in the corresponding horizontal and vertical context. Hence the number of contexts with $\beta \mod 4 = 2$ remains odd.

The action of H_1 subdivides the set of the six contexts into three orbits of size 2. Since the number of non-zero values of β is odd, there must be at least one orbit in which one β has the value 0 and the other has the value 2. Within this orbit, choose $\mathbf{a}, \mathbf{b}$ such that $\beta(\mathbf{a}, \mathbf{b}) = 0$. Hence, $\beta(H_1 \mathbf{a}, H_1 \mathbf{b}) = 2$. Thus, positivity is not preserved under H_1, for any phase convention γ. □

On the other hand, the simulation method of Section IV A has no problem with the above example circuit. Namely, there are Wigner functions of the type Eq. (6) for which the Hadamard gate on a single (the first) qubit is in the group of free gates, $H_1 \in G$, and Z_1 is in the set of directly measurable observables, $Z_1 \in \mathcal{O}$. An example for such a Wigner function is given in Section VI.

We observe that the negativity which can be introduced into a Wigner function by the free unitary gates G is of a very special kind. Namely, it can be lifted by redefinition of the Wigner function according to $A_\mathbf{v} \mapsto A'_\mathbf{v} = g A_\mathbf{v} g^\dagger$, $\forall \mathbf{v} \in V$, for some $g \in G$.

To summarize, while in the present framework the free measurements are required to preserve positivity of the Wigner function, no such constraint needs to be imposed on the free unitaries. The amount of negativity introduced into the Wigner functions by the free unitaries can be large, as measured by mana [38]. However, it is always of a special kind. In this sense, our observation complements the recent finding [39] that a small amount of mana—of any kind—does not compromise the efficiency of a suitable classical simulation algorithm.

V. CONTEXTUALITY

The requirement (P3) on measurements to preserve positivity of the Wigner function has implications for contextuality in QCSI, which we discuss in this section. To begin, we define the notion of "non-contextual hidden variable model" to which our results refer.

A. Non-contextual hidden variable models

Recall that $\mathcal{O}$ is the set of Pauli observables which can be directly measured in QCSI, M is the set of observables which can have their value inferred by measurement of observables in $\mathcal{O}$, and any $C \subset M$ is a set of Pauli observables which can have their value inferred jointly, from a single copy of the given quantum state.

Definition 4 *A non-contextual hidden variable model describing the physical setting $(\rho, \mathcal{O})$ consists of (a) a non-empty set $\mathcal{S}$ of internal states, (b) a probability distribution q over $\mathcal{S}$, and (c) conditional probability distributions $p(\boldsymbol{s}_C|\nu)$, $\nu \in \mathcal{S}$, for outcomes $\boldsymbol{s}_C = (s_1, s_2, .., s_{|C|})$, namely one for each set C of jointly measurable observables, such that*

(i) For every $\nu \in \mathcal{S}$, all observables $O \in M$ have definite values, $\lambda_\nu(O) = \pm 1$, and for all sets C of jointly measurable observables,

$$p(\boldsymbol{s}_C|\nu) = \prod_{i|\, O_i \in C} \delta_{(-1)^{s_i}, \lambda_\nu(O_i)}. \tag{22}$$

(ii) For any triple of jointly measurable observables $A, B, AB \in M$, and all $\nu \in \mathcal{S}$, the value assignments are consistent,

$$\lambda_\nu(AB) = \lambda_\nu(A)\lambda_\nu(B). \tag{23}$$

(iii) Given the quantum state ρ, the probability distribution q_ρ reproduces all probability distributions of measurement outcomes; i.e.

$$p_{C,\rho}(\boldsymbol{s}_C) = \sum_{\nu \in \mathcal{S}} p(\boldsymbol{s}_C|\nu)\, q_\rho(\nu), \tag{24}$$

for all sets C of jointly measurable observables, and all values of $\boldsymbol{s}_C$.

We say that a quantum state ρ is contextual if no non-contextual HVM according to Def. 4 exists that correctly reproduces the probability distributions $p_{C,\rho}(\mathbf{s}_C)$ of measurement outcomes for all sets C of jointly measurable observables.

B. The absence of state-independent contextuality

Recall that, as a consequence of postulate (P1), we have $V_M = V$; cf. Eq. (15). Consider the value assignment

$$\lambda(T_\mathbf{a}) = 1, \;\; \forall \mathbf{a} \in V. \tag{25}$$

First, by Eq. (9), for any three commuting and directly measurable observables $T_\mathbf{a}, T_\mathbf{b}, T_{\mathbf{a}+\mathbf{b}} \in \mathcal{O}$ we have $T_{\mathbf{a}+\mathbf{b}} = +T_\mathbf{a} T_\mathbf{b}$. Thus, the above value assignment is compatible with all available direct measurements.

Second, the value of any observable $T_{\mathbf{a}+\mathbf{b}} \in M\backslash\mathcal{O}$ is inferred by measuring a suitable observable $T_{\mathbf{a}} \in \mathcal{O}$ for which $[T_{\mathbf{a}+\mathbf{b}}, T_{\mathbf{a}}] = 0$, and then running a procedure to infer the value of $T_{\mathbf{b}}$. With Eq. (9), for all such observables $T_{\mathbf{a}}$ it holds that $T_{\mathbf{a}+\mathbf{b}} = +T_{\mathbf{a}}T_{\mathbf{b}}$, and the assignment Eq. (25) is thus consistent.

While Mermin's square an its cousins are present, the operational restriction enforced by condition Eq. (9) prevents obstructions to the assignment Eq. (25) from being established as experimental facts. Hence at least one consistent assignment exists, and there is no state-independent contextuality in this setting.

We are now in the position to prove Lemma 5 of Section III E.

Proof of Lemma 5. If there is a set C with $T_{\mathbf{a}}, T_{\mathbf{b}} \in C$ such then $T_{\mathbf{a}+\mathbf{b}} = -T_{\mathbf{a}}T_{\mathbf{b}}$ then the value assignment $\lambda(T_{\mathbf{a}}) = 1$, $\forall \mathbf{a} \in V$, is inconsistent. Contradiction. □

C. Contextuality implies Wigner negativity

In accordance with existing results [17], [6], also for the present setting of QCSI on qubits it holds that a non-negative Wigner function always implies the viability of a non-contextual hidden variable model.

Theorem 2 *Consider a quantum state ρ with Wigner function W_ρ. If $W_\rho \geq 0$ then the measurement of all Pauli observables $T_{\mathbf{a}} \in \mathcal{O}$ can be described by a non-contextual hidden variable model.*

Proof of Theorem 2. The Wigner function itself constitutes a non-contextual HVM, with set of internal states $\mathcal{S} = V$, probability distribution $q_\rho(\mathbf{u}) = W_\rho(\mathbf{u})$ over the internal states, and the conditional probabilities $p(\mathbf{s}_C|\mathbf{u})$ given by the Wigner functions of the effects.

Using Eq. (7), the probability $p_{\mathbf{a}}(\mathbf{s}_C)$ for obtaining the set $\mathbf{s}_C$ of measurement outcomes corresponding to the effect

$$E_C(\mathbf{s}_C) = \prod_{i|T_{\mathbf{a}(i)}\in C} \frac{I + (-1)^{s_i} T_{\mathbf{a}(i)}}{2},$$

with C a set of simultaneously measurable observables, is

$$p_C(\mathbf{s}_C) = \mathrm{Tr}(\rho E_C(\mathbf{s}_C)) = \sum_{\mathbf{u}\in V} W_\rho(\mathbf{u})(2^n W_{E_C(\mathbf{s}_C)}(\mathbf{u})). \tag{26}$$

If $W_\rho \geq 0$ then W_ρ may be regarded as a probability distribution over the space V of internal states of a hidden variable model. If furthermore $0 \leq 2^n W_{E_C(\mathbf{s}_C)}(\mathbf{u}) \leq 1$ for all $\mathbf{u} \in V$, then we may regard $2^n W_{E_C(\mathbf{s}_C)} =: q_{\mathbf{a}}(\mathbf{s}_C|\mathbf{u})$ as the conditional probability for obtaining the outcome $\mathbf{s}_C$ in the measurement of the observables in C, given the HVM internal state $\mathbf{u} \in V$. Then, Eq. (26) is Bayes' rule for computing the probability $p_{\mathbf{a}}(s)$. We have thus demonstrated item (iii) in Definition 4, provided that $0 \leq 2^n W_{E_C(\mathbf{s}_C)}(\mathbf{u}) \leq 1$ for all $\mathbf{u} \in V$ (see below).

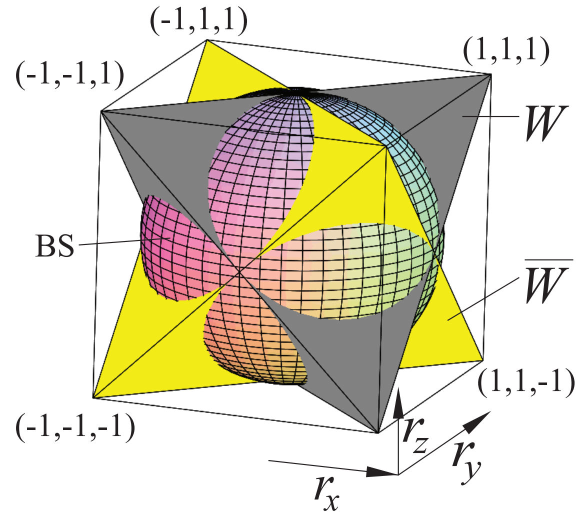


FIG. 3: State space for the one-qubit states $\rho = (I + \mathbf{r}\vec{\sigma})/2$. The physical states lie within or on the Bloch sphere (BS). The two tetrahedra contain the states positively represented by the Wigner functions W and $\overline{W}$, respectively. The state space describable by a non-contextual HVM is a cube with corners $(\pm 1, \pm 1, \pm 1)$; also see [40]. It contains the Bloch ball.

Item (i) of Definition 4. A definite value assignment for all internal states $\mathbf{u} \in V$ has already been established in Eq. (20). It corresponds to

$$2^n W_{E_{\mathbf{a}}(s)}(\mathbf{u}) = \delta_{s,[\mathbf{a},\mathbf{u}]}. \tag{27}$$

for the effects are $E_{\mathbf{a}}(s) = \frac{I+(-1)^s T_{\mathbf{a}}}{2}$, with $s \in \{0, 1\}$.

We now need to consider larger sets C of simultaneously measurable observables, and the corresponding effects $E_C(\mathbf{s}_C) = \prod_{i|T_{\mathbf{a}(i)}\in C} \frac{I+(-1)^{s_i} T_{\mathbf{a}(i)}}{2}$. By an analogous calculation, using Lemma 5, Eq. (22) again follows.

Item (ii) of Definition 4. The final item to check is whether the value assignments Eq. (20) are consistent. Eq. (20) yields $\lambda_{\mathbf{u}}(T_{\mathbf{a}+\mathbf{b}}) = \lambda_{\mathbf{u}}(T_{\mathbf{a}})\lambda_{\mathbf{u}}(T_{\mathbf{b}})$, for all $\mathbf{a}, \mathbf{b} \in V$. By Lemma 5, if $\{T_{\mathbf{a}}, T_{\mathbf{b}}\}$ form a jointly measurable set, then $T_{\mathbf{a}+\mathbf{b}} = T_{\mathbf{a}}T_{\mathbf{b}}$. Hence, $\lambda_{\mathbf{u}}(T_{\mathbf{a}}T_{\mathbf{b}}) = \lambda_{\mathbf{u}}(T_{\mathbf{a}})\lambda_{\mathbf{u}}(T_{\mathbf{b}})$ for such pairs $\mathbf{a}, \mathbf{b} \in V$, as required.

Thus, if $W_\rho \geq 0$ then (V, W_ρ, λ) is a viable non-contextual hidden variable model, with state space V, probability distribution W_ρ over V, and non-contextual value assignment $\lambda_{\mathbf{u}} : V \longrightarrow \mathbb{Z}_2$ given by Eq. (20). □

The converse of Theorem 2 does not hold: there are quantum states with a non-contextual HVM description for which all considered Wigner functions are negative. This is illustrated in Fig. 3 for the example of a single qubit, where all physically allowed states have an HVM description [1]. The one-qubit states are all of the form $\rho = (I + \mathbf{r}\vec{\sigma})/2$, and the physical such states are constrained by $|\mathbf{r}| \leq 1$. The set of states describable in terms of a non-contextual HVM is a cube, $|r_x|, |r_y|, |r_z| \leq 1$, containing all physical states. The eight extremal states i of this cube have definite value assignments $\lambda_i(X), \lambda_i(Y), \lambda_i(Z) = \pm 1$ for the observables X, Y, Z.

Up to equivalence under translation, there are two one-qubit Wigner functions of type Eq. (6), namely the Wigner function W defined by the phase point operator at the origin $A_{\mathbf{0}} = (I + X + Z + Y)/2$, and the Wigner function $\overline{W}$ defined by $\overline{A}_{\mathbf{0}} = HA_{\mathbf{0}}H^\dagger = (I+X+Z-Y)/2$. The phase space for these Wigner functions is $\mathbb{Z}_2 \times \mathbb{Z}_2$, and W, $\overline{W}$ thus four extremal states each. If these extremal states are combined, the extremal states of the non-contextual HVM are recovered. Each Wigner function by itself has only half of the extremal states of the HVM, and the set of positively represented states is thus smaller. Furthermore, there are physical states which are negatively represented by both W and $\overline{W}$; See Fig. 3. Contextuality and negativity of the Wigner functions Eq. (6) are thus not the same.

Remark: While there are physical one-qubit quantum states which are negatively represented by both W and $\overline{W}$, every state $\rho = (I + \mathbf{r}\vec{\sigma})/2$, with $|r_x|, |r_y|, |r_z| \leq 1$, can be described by an ensemble

$$\mathcal{E}(\rho) = \{(p_1, \rho_1), (p_2, \rho_2)\},$$

such that $W_{\rho_1} \geq 0$ and $\overline{W}_{\rho_2} \geq 0$. A generalization of this fact to n-qubit systems will be of relevance in Section V D.

D. Contextuality as a resource

Theorem 3 *For any QCSI scheme $(\mathcal{O}, G, \Omega, \mathcal{M})$, if the input magic state ρ_{in} can be described by a non-contextual HVM, then the quantum state $\rho_t(\boldsymbol{s}_{\prec t})$ at time t, conditioned on the prior measurement record $\boldsymbol{s}_{\prec t}$, can be described by a non-contextual HVM, for any t and any $\boldsymbol{s}$.*

Thus, a QCSI scheme $(\mathcal{O}, G, \Omega, \mathcal{M})$ can only serve as a universal state preparator if it has access to contextual magic states. In this sense, contextuality is a resource for all QCSI schemes $(\mathcal{O}, G, \Omega, \mathcal{M})$.

Before we give the proof of Theorem 3, we need to set up some more notation. We observe that the state space of a general non-contextual HVM is larger than the state space of an HVM deriving from a non-negative Wigner function; See the discussion of a single qubit in Section V C/ Fig. 3. The enlarged state space $\mathcal{S} = \{\nu\}$ is finite, yet maximal in the sense that, for every value assignment $\lambda(\cdot)$ satisfying the consistency condions Eq. (23), there is a corresponding internal state $\nu \in \mathcal{S}$ such that $\lambda_\nu(\cdot) \equiv \lambda(\cdot)$.

We choose to have this state space $\mathcal{S}$ acted upon by the group V, dividing $\mathcal{S}$ into orbits. Namely, given an element $\nu \in \mathcal{S}$ specified by the value assignment $\lambda_\nu : V \longrightarrow \{\pm 1\}$, there is another internal state $\nu + \mathbf{u}$, defined through the value assignment

$$\lambda_{\nu+\mathbf{u}}(\mathbf{a}) = \lambda_\nu(\mathbf{a})(-1)^{[\mathbf{u},\mathbf{a}]},\ \forall \mathbf{a} \in V, \tag{28}$$

for all $\mathbf{u} \in V$. In Eq. (28), we have set $\lambda_\nu(\mathbf{a}) := \lambda_\nu(T_{\mathbf{a}})$ for notational simplicity. It is easily seen that $\nu \in \mathcal{S} \Leftrightarrow \nu + \mathbf{u} \in \mathcal{S}$, for all $\mathbf{u} \in V$. The condition to check is the consistency of the value assignment in item (ii) of Def. 4. Eq. (23) is preserved under the change $\lambda_\nu(\mathbf{a}) \mapsto \lambda_\nu(\mathbf{a})(-1)^{[\mathbf{u},\mathbf{a}]}$, for any $\mathbf{u} \in V$. The group action of V on $\mathcal{S}$ defined through Eq. (28) labels the elements of $\mathcal{S}$ in a fashion convenient for the subsequent discussion.

Proof of Theorem 3. An HVM with a set $\mathcal{S}$ of internal states and probability distribution q over $\mathcal{S}$ correctly describes a quantum state ρ if

$$\langle T_{\mathbf{a}} \rangle_\rho = \sum_{\nu \in \mathcal{S}} q(\nu)\lambda_\nu(\mathbf{a}),\ \forall \mathbf{a} \in V. \tag{29}$$

This holds because the $T_{\mathbf{a}}, \mathbf{a} \in V$, form an operator basis for the Hilbert space of n qubits.

The proof of Theorem 3 is by induction. We assume that there exists an HVM with probability distribution $q_{t,\mathbf{s}_{\prec t}}$ which describes the quantum state $\rho_t(\mathbf{s}_{\prec t})$, conditioned on the previous measurement record $\mathbf{s}_{\prec t}$. We then show that there is an HVM with probability distribution $q_{t+1,\mathbf{s}_{\prec t+1}}$ which describes the quantum state $\rho_{t+1}(\mathbf{s}_{\prec t+1})$.

To establish this result, we need the relation between $q_{t+1,\mathbf{s}_{\prec t+1}}$ and its precursor $q_{t,\mathbf{s}_{\prec t}}$. Denoting the observable measured in the t-th time step of the computation by $T_{\mathbf{a}_t} \in \mathcal{O}$ and the corresponding measurement outcome by $s_t \in \mathbb{Z}_2$, the required relation is

$$q_{t+1,\mathbf{s}_{\prec t+1}}(\nu) = \frac{\delta_{(-1)^{s_t},\lambda_\nu(\mathbf{a}_t)}}{\overline{p}_t(s_t|\mathbf{s}_{\prec t})} \frac{q_{t,\mathbf{s}_{\prec t}}(\nu) + q_{t,\mathbf{s}_{\prec t}}(\nu + \mathbf{a}_t)}{2}, \tag{30a}$$

$$\overline{p}_t(s_t|\mathbf{s}_{\prec t}) = \sum_{\nu \in \mathcal{S}} \delta_{(-1)^{s_t},\lambda_\nu(\mathbf{a}_t)} q_{t,\mathbf{s}_{\prec t}}(\nu). \tag{30b}$$

Therein, $\overline{p}_t(s_t|\mathbf{s}_{\prec t})$ is the HVM prediction for the probability of obtaining the outcome s_t in the measurement of $T_{\mathbf{a}_t}$, given a prior measurement record $\mathbf{s}_{\prec t}$. Eq. (30) will be justified a posteriori. Namely, with these assignments, the induction argument works out.

With Eq. (29), the induction assumption is

$$\langle T_{\mathbf{a}} \rangle_{\rho_t} = \langle T_{\mathbf{a}} \rangle_{q_t},\ \forall \mathbf{a} \in V.$$

Therein, we have supressed the dependence on the measurement record, to simplify the notation. We need to show that

$$\langle T_{\mathbf{a}} \rangle_{\rho_{t+1}} = \langle T_{\mathbf{a}} \rangle_{q_{t+1}},\ \forall \mathbf{a} \in V,$$

and that $\overline{p}_t(s_t|\mathbf{s}_{\prec t}) = p_t(s_t|\mathbf{s}_{\prec t})$, with $p_t(s_t|\mathbf{s}_{\prec t})$ the quantum mechanical value for the probability of the outcome s_t given the prior measurement record $\mathbf{s}_{\prec t}$.

First, regarding the probability of finding s_t,

$$\begin{aligned}
\overline{p}_t(s_t|\mathbf{s}_{\prec t}) &= \sum_{\nu\in\mathcal{S}} \frac{1+(-1)^{s_t}\lambda_\nu(\mathbf{a}_t)}{2} q_{t,\mathbf{s}_{\prec t}}(\nu) \\
&= \frac{1}{2}\sum_{\nu\in\mathcal{S}} q_{t,\mathbf{s}_{\prec t}}(\nu) + \frac{(-1)^{s_t}}{2}\sum_{\nu\in\mathcal{S}} q_{t,\mathbf{s}_{\prec t}}(\nu)\lambda_\nu(\mathbf{a}_t) \\
&= \frac{\langle I\rangle_{\rho_t(\mathbf{s}_{\prec t})} + (-1)^{s_t}\langle T_{\mathbf{a}_t}\rangle_{\rho_t(\mathbf{s}_{\prec t})}}{2} \\
&= \mathrm{Tr}\left(\rho_t(\mathbf{s}_{\prec t})\frac{I+(-1)^{s_t}T_{\mathbf{a}_t}}{2}\right) \\
&= p_t(s_t|\mathbf{s}_{\prec t}).
\end{aligned}$$

We thus reproduce the quantum mechanical expression within the HVM. Above, in transitioning from the second to the third line we have invoked the induction assumption.

Second, regarding the expectation values of the $T_\mathbf{a}$ on $\rho_{t+1}(\mathbf{s}_{\prec t+1})$, the HVM prediction is

$$\begin{aligned}
\langle T_\mathbf{a}\rangle_{q_{t+1}} &= \sum_{\nu\in\mathcal{S}} q_{t+1,\mathbf{s}_{\prec t+1}}(\nu)\lambda_\nu(\mathbf{a}) \\
&= \sum_{\nu\in\mathcal{S}} \frac{1+(-1)^{s_t}\lambda_\nu(\mathbf{a}_t)}{4p_t(s_t|\mathbf{s}_{\prec t})} q_{t,\mathbf{s}_{\prec t}}(\nu)\lambda_\nu(\mathbf{a}) + \\
&\quad + \sum_{\nu\in\mathcal{S}} \frac{1+(-1)^{s_t}\lambda_\nu(\mathbf{a}_t)}{4p_t(s_t|\mathbf{s}_{\prec t})} q_{t,\mathbf{s}_{\prec t}}(\nu+\mathbf{a}_t)\lambda_\nu(\mathbf{a}).
\end{aligned}$$

Reordering the sum via the substitution $\nu+\mathbf{a}_t\rightarrow\nu$, and using Eq. (28), the second term in the last line equals

$$\sum_{\nu\in\mathcal{S}} \frac{1+(-1)^{s_t}\lambda_\nu(\mathbf{a}_t)}{4p_t(s_t|\mathbf{s}_{\prec t})} q_{t,\mathbf{s}_{\prec t}}(\nu)\lambda_\nu(\mathbf{a})(-1)^{[\mathbf{a},\mathbf{a}_t]}.$$

We now distinguish between the case where $T_\mathbf{a}, T_{\mathbf{a}_t}$ commute and where they don't.

Case (i): $[\mathbf{a},\mathbf{a}_t]=1$. Then, $\langle T_\mathbf{a}\rangle_{q_{t+1}}=0$, which is the correct quantum mechanical expression.

Case (ii): $[\mathbf{a},\mathbf{a}_t]=0$. Then, the expression for $\langle T_\mathbf{a}\rangle_{q_{t+1}}$ simplifies to

$$\begin{aligned}
\langle T_\mathbf{a}\rangle_{q_{t+1}} &= \sum_{\nu\in\mathcal{S}} \frac{1+(-1)^{s_t}\lambda_\nu(\mathbf{a}_t)}{2p_t(s_t|\mathbf{s}_{\prec t})} q_{t,\mathbf{s}_{\prec t}}(\nu)\lambda_\nu(\mathbf{a}) \\
&= \frac{1}{2p_t(s_t|\mathbf{s}_{\prec t})}\sum_{\nu\in\mathcal{S}} q_{t,\mathbf{s}_{\prec t}}(\nu)\lambda_\nu(\mathbf{a}) + \\
&\quad + \frac{(-1)^{s_t}}{2p_t(s_t|\mathbf{s}_{\prec t})}\sum_{\nu\in\mathcal{S}} q_{t,\mathbf{s}_{\prec t}}(\nu)\lambda_\nu(\mathbf{a}+\mathbf{a}_t).
\end{aligned}$$

Here we have used the relation $\lambda_\nu(\mathbf{a}+\mathbf{a}_t)=\lambda_\nu(\mathbf{a}_t)\lambda_\nu(\mathbf{a})$, which arises as follows. Since $T_{\mathbf{a}_t}\in\mathcal{O}$, $T_\mathbf{a}\in M$, and $[T_{\mathbf{a}_t},T_\mathbf{a}]=0$ by the case assumption, $\{T_{\mathbf{a}_t},T_\mathbf{a}\}$ is a jointly measurable set of observables; cf. example (iii) after Def. 3. (The procedure is to measure $T_{\mathbf{a}_t}\in\mathcal{O}$ first, and then run the measurement sequence for $T_\mathbf{a}\in M$.) Thus, by Property (ii) of Def. 4 for non-contextual HVMs, $\lambda_\nu(T_{\mathbf{a}_t}T_\mathbf{a})=\lambda_\nu(T_{\mathbf{a}_t})\lambda_\nu(T_\mathbf{a})$. Finally, with Lemma 1, $T_{\mathbf{a}_t}T_\mathbf{a}=T_{\mathbf{a}+\mathbf{a}_t}$, which yields the stated relation.

Next we use the induction assumption, and obtain

$$\begin{aligned}
\langle T_\mathbf{a}\rangle_{q_{t+1}} &= \frac{1}{2p_t(s_t|\mathbf{s}_{\prec t})}\left(\langle T_\mathbf{a}\rangle_{\rho_t}(-1)^{s_t}+\langle T_{\mathbf{a}+\mathbf{a}_t}\rangle_{\rho_t}\right) \\
&= \frac{1}{p_t(s_t|\mathbf{s}_{\prec t})}\mathrm{Tr}\left(\rho_t\frac{I+(-1)^{s_t}T_{\mathbf{a}_t}}{2}T_\mathbf{a}\right) \\
&= \frac{\mathrm{Tr}\left(\left[\frac{I+(-1)^{s_t}T_{\mathbf{a}_t}}{2}\rho_t\frac{I+(-1)^{s_t}T_{\mathbf{a}_t}}{2}\right]T_\mathbf{a}\right)}{p_t(s_t|\mathbf{s}_{\prec t})} \\
&= \langle T_\mathbf{a}\rangle_{\rho_{t+1}}.
\end{aligned}$$

We thus reproduce the quantum mechanical expression within the HVM. This completes the induction step.

The induction starts at time $t=1$, where $\rho_1=\rho_{\mathrm{in}}$ has an HVM description, by assumption of Theorem 3. Thus, by induction, for every time $t\geq 1$ and every history $\mathbf{s}_{\prec t}$ of measurement outcomes, the conditional state $\rho_t(\mathbf{s}_{\prec t})$ has a description in terms of a non-contextual HVM. □

Corollary 1 *For any QCSI scheme $(\mathcal{O},G,\Omega,\mathcal{M})$, if the input magic state ρ_{in} can be described by a non-contextual HVM, then for the measurement of any sequence of observables $\{T_{\mathbf{a}_t}, t=1..t_{max}\}\subset\mathcal{O}$, the probability distribution $p(\mathbf{s})=p(s_1,s_2,..,s_{t_{max}})$ of outcomes is fixed by the HVM for ρ_{in}. The $T_{\mathbf{a}_t}$ may be mutually non-commuting and dependent on previous measurement outcomes.*

Proof of Corollary 1. By Bayes' rule, the joint probability of the outcomes $\mathbf{s}$ can be written as

$$p(\mathbf{s}) = \prod_{t=1}^{t_{\max}} p_t(s_t|\mathbf{s}_{\prec t}).$$

By Theorem 3, the conditional probabilities $p_t(s_t|\mathbf{s}_{\prec t})=\overline{p}_t(s_t|\mathbf{s}_{\prec t})$ are all correctly obtained from the probability distributions $q_{t,\mathbf{s}_{\prec t}}$, cf. Eq. (30b). The distributions $q_{t,\mathbf{s}_{\prec t}}$, for $t=2,..,t_{\max}$, in turn follow from the distribution $q_{1,\mathbf{s}_{\prec 1}=\emptyset}$ (describing ρ_{in} at $t=1$), by Eq. (30a). Thus, $p(\mathbf{s})$ is fully specified by the probability distribution $q_{1,\mathbf{s}_{\prec 1}=\emptyset}$ over the state space $\mathcal{S}$ of the HVM. □

E. Generalized simulation algorithm

Theorem 4 *For any QCSI scheme $(\mathcal{O},G,\Omega,\mathcal{M})$, if (i) the input magic state ρ_{in} can be described by a non-contextual HVM with state space $\mathcal{S}$ and value assignments $\lambda_\nu: V\longrightarrow\{\pm1\}$, for all $\nu\in\mathcal{S}$, (ii) this HVM can be efficiently sampled from, and (iii) the value assignments $\lambda_\nu(\boldsymbol{a})$ and the phase convention $\gamma(\boldsymbol{a})$ can be efficiently evaluated for all $\boldsymbol{a}\in V_\mathcal{O}$, then any resulting QCSI can be efficiently classically simulated.*

Theorem 4 is proved constructively, i.e., by providing a classical simulation algorithm. This algorithm is

Algorithm 2

1. Draw a sample $\nu \in \mathcal{S}$ from the probability distribution $q_{1,\mathbf{s}_{\prec 1}=\emptyset}$ describing ρ_{in} in the HVM, and set $\nu_1 := \nu$.

2. For all the measurements of observables $T_{\mathbf{a}_t} \in \mathcal{O}$ comprising the circuit, starting with the first,

   (a) Output the measurement outcome $\lambda_{\nu_t}(\mathbf{a}_t) \in \{\pm 1\}$ for the observable $T_{\mathbf{a}_i}$,

   (b) Flip a fair coin, and update the sample
$$\nu_t \longrightarrow \nu_{t+1} = \begin{cases} \nu_t, & \text{if “heads”} \\ \nu_t + \mathbf{a}_t, & \text{if “tails'} \end{cases}, \tag{31}$$
   until the measurement sequence is complete.

3. Repeat until sufficient statistics is gathered.

This algorithm is an almost exact copy of the simulation algorithm encountered in Section IV B, and we comment on the resemblance in Section V F.

Before we proceed to the proof of Theorem 4, we briefly discuss what sampling from conditional probability distributions means for the above algorithm. For any sample ν drawn in Step 1, while looping through Step 2, a measurement record $\mathbf{s}$ is built up. In every iteration t of Step 2, the updated sample ν_t may be regarded as being drawn from a probability distribution $\tilde{q}_{t,\mathbf{s}_{\prec t}}$, conditioned on the previous measurement record $\mathbf{s}_{\prec t}$. So the above simulation algorithm definitely samples. The question is whether it samples from the *correct* distributions, i.e., whether $\tilde{q}_{t,\mathbf{s}_{\prec t}} = q_{t,\mathbf{s}_{\prec t}}$, for all $t = 1, .., t_{\max}$ and for all $\mathbf{s}$.

Proof of Theorem 4. The proof proceeds by demonstrating the correctness and efficiency of the above classical simulation algorithm.

Correctness. We first show that for each time t and measurement record $\mathbf{s}_{\prec t}$, the above classical simulation algorithm (i) produces the correct quantum-mechanical conditional probability $p_t(s_t|\mathbf{s}_{\prec t})$ of obtaining the outcome s_t in the measurement of the observable $T_{\mathbf{a}_t} \in \mathcal{O}$, and (ii) samples from the correct conditional probability distribution $q_{t,\mathbf{s}_{\prec t}}$ of the HVM, which is given by Eq. (30).

The proof is by induction. We assume that at time t, the classical simulation algorithm samples from the correct distribution $q_{t,\mathbf{s}_{\prec t}}$.

Re (i): Denote the conditional probabilities produced by the simulation algorithm as $\tilde{p}_t(s_t|\mathbf{s}_{\prec t})$. A state $\nu \in \mathcal{S}$ contributes its probability weight $q_{t,\mathbf{s}_{\prec t}}(\nu)$ to $\tilde{p}_t(0|\mathbf{s}_{\prec t})$ or $\tilde{p}_t(1|\mathbf{s}_{\prec t})$ if $\lambda_\nu(T_{\mathbf{a}_t}) = +1$ or $\lambda_\nu(T_{\mathbf{a}_t}) = -1$, respectively. Therefore,
$$\tilde{p}_t(s_t|\mathbf{s}_{\prec t}) = \sum_{\nu \in \mathcal{S}} \delta_{\lambda_\nu(T_{\mathbf{a}_t}),(-1)^{s_t}} q_{t,\mathbf{s}_{\prec t}}(\nu) = \overline{p}_t(s_t|\mathbf{s}_{\prec t}).$$
The second equality follows by comparison with Eq. (30b). Furthermore, $\overline{p}_t(s_t|\mathbf{s}_{\prec t}) = p_t(s_t|\mathbf{s}_{\prec t})$ was already demonstrated in the proof of Theorem 3. Thus, $\tilde{p}_t(s_t|\mathbf{s}_{\prec t}) = p_t(s_t|\mathbf{s}_{\prec t})$, as required.

Re (ii): Through the value assignment in step 2(a), an internal state $\nu_t \in \mathcal{S}$ contributes to
$$\begin{array}{ll} \tilde{q}_{t+1,(\mathbf{s}_{\prec t},s_t=0)}, & \text{if } \lambda_{\nu_t}(\mathbf{a}_t) = +1, \\ \tilde{q}_{t+1,(\mathbf{s}_{\prec t},s_t=1)}, & \text{if } \lambda_{\nu_t}(\mathbf{a}_t) = -1. \end{array}$$
The update rule for Step 2(a) is thus
$$q_{t,\mathbf{s}_{\prec t}}(\tau) \overset{s_t}{\longrightarrow} q'_{t+1,\mathbf{s}_{\prec t+1}}(\tau) = q_{t,\mathbf{s}_{\prec t}}(\tau) \frac{\delta_{\lambda_\tau(\mathbf{a}_t),(-1)^{s_t}}}{p_t(s_t|\mathbf{s}_{\prec t})},$$
for all $\tau \in \mathcal{S}$, and $p_t(s_t|\mathbf{s}_{\prec t})$ appears for normalization.

In step 2(b), with Eq. (31), the update rule is
$$q'_{t+1,\mathbf{s}_{\prec t+1}} \longrightarrow \tilde{q}_{t+1,\mathbf{s}_{\prec t+1}} = q'_{t+1,\mathbf{s}_{\prec t+1}} * \frac{\delta_{\mathbf{0}} + \delta_{\mathbf{a}_t}}{2},$$
where “$*$” stands for convolution. Using Eq. (28), the resulting expression for $\tilde{q}_{t+1,\mathbf{s}_{\prec t+1}}(\nu)$ matches the expression in Eq. (30a), i.e., $\tilde{q}_{t+1,\mathbf{s}_{\prec t+1}}(\nu) = q_{t+1,\mathbf{s}_{\prec t+1}}(\nu)$, as required. This completes the induction step.

The induction assumption is satisfied for $t = 1$, by the first assumption of Theorem 4. Thus, by induction, the above algorithm samples from the correct conditional outcome probabilities $p(s_t|\mathbf{s}_{\prec t})$ for measurement outcomes s_t and from the correct HVM distributions $q_{t,\mathbf{s}_{\prec t}}$ describing $\rho_t(\mathbf{s}_{\prec t})$, for all times t and all outcome histories $\mathbf{s}$.

Efficiency. The classical preprocessing of removing the unitaries $g \in G$ from the circuit is efficient if the function $\gamma : V \longrightarrow \mathbb{Z}_4$ can be efficiently computed, which holds by assumption. See the proof of Theorem 1.

Regarding the simulation algorithm itself, the critical step is 2(a), the evaluation of the function λ_ν on some $\mathbf{a} \in V$. Again, the efficiency of this function evaluation is guaranteed by the assumption of the theorem. □

F. Relation between Algorithms 1 and 2

Algorithms 1 and 2 are very similar. They only differ in the sampling source they have access to. By Theorem 2, the sampling source for Algorithm 2, based on non-contextual HVMs, is at least as powerful as the sampling source for Algorithm 1, based on non-negative Wigner functions. By the 1-qubit example of Section V C, the former source is indeed more powerful.

In the following, we investigate the relation between the two sampling sources in greater detail. The result is that if the initial quantum state ρ_{in} can be described by a non-contextual HVM, then it can be represented by an ensemble
$$\mathcal{E}_{\rho_{\text{in}}} = \{(p_i, \rho_i)\},$$

such that there are Wigner functions W^{γ_i} for which (i) $W^{\gamma_i}_{\rho_i} \geq 0$, $\forall i$, and (ii) the measurement of observables from the set $\mathcal{O}$ preserves positivity of the W^{γ_i}, $\forall i$.

Therefore, Algorithm 2 can be simulated by a master algorithm that merely draws samples $\nu \in \mathcal{S}$ from the non-contextual HVM, and then employs Algorithm 1 as a subroutine for dealing with the samples. This interpretation of Algorithm 2 is developed below.

The state space $\mathcal{S}$ of the HVM can be partitioned into orbits $[\nu]$ of V,

$$[\nu] = \{\nu + \mathbf{u}, \mathbf{u} \in V\} \in \mathcal{S}/V.$$

Then there exists a special orbit $[0] \in \mathcal{S}/V$ defined by the property that there is a $\tau_{[0]} \in [0]$ for which the value assignment is constant, $\lambda_{\tau_{[0]}}(\cdot) \equiv 1$. With Eq. (28) it then follows that

$$\lambda_{\tau_{[0]}+\mathbf{u}}(\mathbf{a}) = (-1)^{[\mathbf{u},\mathbf{a}]},\ \forall \mathbf{a} \in V.$$

Comparing with Eq. (20), we find that the above value assignment $\lambda_{\tau_{[0]}+\mathbf{u}}(\mathbf{a})$ agrees with the value assignment made by a positive Wigner function Eq. (6) considered as an HVM, if we identify, for all $\mathbf{u} \in V$,

$$(\tau_{[0]} + \mathbf{u}) \in [0] \longleftrightarrow \mathbf{u} \in V.$$

This motivates the definition of a quantum state $\rho_{[0]}$ associated with the orbit $[0]$, via

$$W_{\rho_{[0]}}(\mathbf{u}) := \frac{q(\tau_{[0]} + \mathbf{u})}{p_{[0]}},\ \forall \mathbf{u} \in V, \tag{32}$$

where $p_{[0]} = \sum_{\mathbf{u}\in V} q(\tau_{[0]}+\mathbf{u})$ to ensure proper normalization. The state $\rho_{[0]}(\mathbf{u})$ is not required to be a valid quantum state, i.e., to be positive semi-definite. The only requirement is a non-negative Wigner function, which is satisfied by definition. The fact that classical sampling algorithms can handle states which have a positive Wigner function but are not proper quantum states is familiar from the qudit case [9].

In analogy with Eq. (32), we will now define states $\rho_{[\nu]}$ for all orbits $[\nu] \in \mathcal{S}/V$. The state $\rho_{[0]}$ and its cousins will then be used in the interpretation of Algorithm 2.

For any $[\nu] \in \mathcal{S}/V$, pick a $\tau_{[\nu]} \in [\nu]$ and define

$$T_{\mathbf{a}}^{\gamma_{[\nu]}} := \lambda^{-1}_{\tau_{[\nu]}}(\mathbf{a}) T_{\mathbf{a}},\ \forall \mathbf{a} \in V, \tag{33}$$

where on the r.h.s. $T_{\mathbf{a}} = T^{\gamma}_{\mathbf{a}}$, as defined in Eq. (5). Denoting $\lambda_{\tau_{[\nu]}}(\mathbf{a}) = (-1)^{s_{[\nu]}(\mathbf{a})}$, for all $\mathbf{a} \in V$, we thus have the relation

$$\gamma_{[\nu]} \equiv \gamma + 2s_{[\nu]} \mod 4. \tag{34}$$

From the above definition of $T_{\mathbf{a}}^{\gamma_{[\nu]}}$, $\lambda_{\tau_{[\nu]}}\left(T_{\mathbf{a}}^{\gamma_{[\nu]}}\right) = 1$, for all $\mathbf{a} \in V$. We can thus reproduce for any orbit $[\nu]$ the previous argument for $[0]$. First, with Eq. (28),

$$\lambda_{\tau_{[\nu]}+\mathbf{u}}\left(T_{\mathbf{a}}^{\gamma_{[\nu]}}\right) = (-1)^{[\mathbf{a},\mathbf{u}]},\ \forall \mathbf{a}, \mathbf{u} \in V. \tag{35}$$

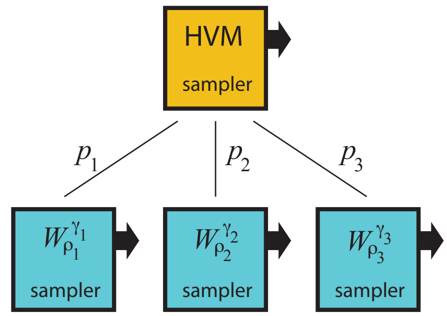


FIG. 4: Relation between Algorithms 1 and 2. Sampling from the probability distribution underlying a non-contextual HVM may be viewed as a two-stage process. Stage 1: Sampling from a probability distribution $\{p_i\}$ over Wigner functions, Stage 2: Sampling from the phase space w.r.t. to the Wigner function chosen in the first stage.

Again by comparison with Eq. (20), the value assignments by the HVM and by the Wigner function $W^{\gamma_{[\nu]}}$ match if we identify, for all $\mathbf{u} \in V$,

$$(\tau_{[\nu]} + \mathbf{u}) \in [\nu] \longleftrightarrow \mathbf{u} \in V. \tag{36}$$

A state $\rho_{[\nu]}$ associated with any orbit $[\nu] \in \mathcal{S}/V$ can now be defined, via

$$W^{\gamma_{[\nu]}}_{\rho_{[\nu]}}(\mathbf{u}) := \frac{q(\tau_{[\nu]} + \mathbf{u})}{p_{[\nu]}},\ \forall \mathbf{u} \in V. \tag{37}$$

Therein, $p_{[\nu]} = \sum_{\mathbf{u}\in V} q(\tau_{[\nu]} + \mathbf{u})$. As before with $\rho_{[0]}(\mathbf{u})$, the state $\rho_{[\nu]}(\mathbf{u})$ is not required to be positive semi-definite.

Remark: For each $[\nu] \in \mathcal{S}/V$, the choice of the representative $\tau_{[\nu]}$ is arbitrary. Different choices lead to different $\gamma_{[\nu]}$, which are, however, related in a simple way. Namely, the corresponding Wigner functions differ only by translation. By contrast, the Wigner functions $W^{\gamma_{[\nu]}}$ and $W^{\gamma_{[\nu']}}$, for any $[\nu'] \neq [\nu]$, are not equivalent under translation.

With the above, we can now re-interpret the sampling from the HVM as the following two-stage process. In the first stage, equivalence classes $[\nu] \in \mathcal{S}/V$ are sampled from, according to the probabilities $\{p_{[\nu]}\}$. In the second stage, given a particular class $[\nu]$, the phase space V is sampled from, according to the conditional probability distribution $q|_{[\nu]}/p_{[\nu]}$. The conditional probability distributions $q|_{[\nu]}/p_{[\nu]}$ over V are regarded as Wigner functions $W^{\gamma_{[\nu]}}_{\rho_{[\nu]}}$ of states $\rho_{[\nu]}$ associated with the orbits $[\nu]$, cf. Eq. (37). See Fig. 4 for illustration.

Algorithm 2 can thus be simulated by a master algorithm calling Algorithm 1 as a subroutine, as follows. Step 1: A sample $\nu \in \mathcal{S}$ is drawn. This sample is converted into the into the pair $([\nu] \in \mathcal{S}, \mathbf{u} \in V)$, such that $\nu = \tau_{[\nu]} + \mathbf{u}$. Step 2: Algorithm 1 is called, with the sample $\mathbf{u}_1 := \mathbf{u}$ being handed over.

The orbit $[\nu]$ has no influence on how Algorithm 1 runs, but needs to be taken into account when the simulated measurement outcomes are returned. Namely, Algorithm 1 returns the values for $T_{\mathbf{a}_t}^{\gamma_{[\nu]}}$, not for the $T_{\mathbf{a}_t}$ with the standard phase convention γ. A conversion of those values is thus necessary, which proceeds by Eq. (33).

There is one more item to check: To employ sampling from Wigner functions $W^{\gamma_{[\nu]}}$ as a subroutine, the measurement of observables which leave positive Wigner functions W positive must also leave all Wigner functions $W^{\gamma_{[\nu]}}$ positive. Denote by $\mathcal{O}_{[\nu]}$ the non-extendable set of directly measurable observables w.r.t. the phase convention $\gamma_{[\nu]}$ (i.e., the set of Pauli observables whose measurement preserves non-negativity of the Wigner function $W^{\gamma_{[\nu]}}$). By Lemma 4, $\mathcal{O}_{[\nu]}$ is unique and of form Eq. (16). Then, we have the following result.

Lemma 8 *For all $[\nu] \in \mathcal{S}/V$, it holds that $\mathcal{O}_{[\nu]} = \mathcal{O}_{max}$.*

In addition, we note that the states $\rho_{[\nu]}$, defined in Eq. (37) have the following relation to the input state ρ_{in} of the computation.

Lemma 9 *For any QCSI scheme $(\mathcal{O}, G, \Omega, \mathcal{M})$, if the input state ρ_{in} has a non-contextual HVM description, then the states $\rho_{[\nu]}$ provide an ensemble representation $\mathcal{E}_{\rho_{in}} = \left\{(p_{[\nu]}, \rho_{[\nu]}), [\nu] \in \mathcal{S}/V\right\}$ of ρ_{in}, i.e.,*

$$\rho_{in} = \sum_{[\nu]\in\mathcal{S}/V} p_{[\nu]}\rho_{[\nu]}. \tag{38}$$

Multiple Wigner functions have previously been discussed in the context of QCSI [10]. Therein, a quantum state is considered classical if *all* its Wigner functions are positive. Our viewpoint is the opposite. For a state to be considered classical, not even a single one of its Wigner functions has to be positive.

Proof of Lemma 8. By Eq. (8), for any phase convention γ it holds that

$$\beta(\mathbf{a},\mathbf{b}) = \gamma(\mathbf{a}) + \gamma(\mathbf{b}) - \gamma(\mathbf{a}+\mathbf{b}) + 2\mathbf{a}_X\mathbf{b}_Z \mod 4.$$

Then, by Eq. (34), the function β based on a specific γ and the functions $\beta_{[\nu]}$ based on the corresponding $\gamma_{[\nu]}$ are related via

$$\beta_{[\nu]}(\mathbf{a},\mathbf{b}) = \beta(\mathbf{a},\mathbf{b}) + 2\left(s_{[\nu]}(\mathbf{a}) + s_{[\nu]}(\mathbf{b}) - s_{[\nu]}(\mathbf{a}+\mathbf{b})\right),$$

where the addition is again mod 4. Now assume that $T_\mathbf{a} \in \mathcal{O}$ and that $[\mathbf{a},\mathbf{b}] = 0$. Then, $\{T_\mathbf{a}, T_\mathbf{b}\}$ is a jointly measurable set of observables, and thus, by Property (ii) of Def. 4, $s_{[\nu]}(\mathbf{a}) + s_{[\nu]}(\mathbf{b}) - s_{[\nu]}(\mathbf{a}+\mathbf{b}) \mod 2 = 0$. Hence,

$$\beta_{[\nu]}(\mathbf{a},\mathbf{b}) = \beta(\mathbf{a},\mathbf{b}),\ \forall\,[\nu] \in \mathcal{S}/V,$$

for all above pairs $\mathbf{a} \in V_\mathcal{O}$, $\mathbf{b} \in V$. Thus, by Lemma 1, the measurement of $T_\mathbf{a}$ preserves positivity of the Wigner function W^γ if and only if it preserves positivity of the Wigner function $W^{\gamma_{[\nu]}}$, for any $[\nu] \in \mathcal{S}/V$. $\square$

Proof of Lemma 9. Since the $T_\mathbf{a}$, $\mathbf{a} \in V$, form a basis of Hermitian operators on n qubits, it suffices to show that $\langle T_\mathbf{a}\rangle_{\rho_{\text{in}}} = \langle T_\mathbf{a}\rangle_{\sum_{[\nu]} p_{[\nu]}\rho_{[\nu]}}$, for all $\mathbf{a} \in V$.

$$\begin{aligned}
\langle T_\mathbf{a}\rangle_{\sum_{[\nu]} p_{[\nu]}\rho_{[\nu]}} &= \sum_{[\nu]} p_{[\nu]}\langle T_\mathbf{a}\rangle_{\rho_{[\nu]}}\\
&= \sum_{[\nu]} p_{[\nu]}\lambda_{\tau_{[\nu]}}(\mathbf{a})\langle T_\mathbf{a}^{\gamma_{[\nu]}}\rangle_{\rho_{[\nu]}}\\
&= \sum_{[\nu]} p_{[\nu]}\lambda_{\tau_{[\nu]}}(\mathbf{a})2^n\sum_{\mathbf{u}\in V} W_{\rho_{[\nu]}}^{\gamma_{[\nu]}}(\mathbf{u})W_{T_\mathbf{a}^{\gamma_{[\nu]}}}^{\gamma_{[\nu]}}(\mathbf{u})\\
&= \sum_{[\nu]} \lambda_{\tau_{[\nu]}}(\mathbf{a})\sum_{\mathbf{u}\in V} q(\tau_{[\nu]}+\mathbf{u})(-1)^{[\mathbf{a},\mathbf{u}]}\\
&= \sum_{\nu\in\mathcal{S}} q(\nu)\lambda_{\tau_{[\nu]}}(\mathbf{a})\lambda_\nu\left(T_\mathbf{a}^{\gamma_{[\nu]}}\right)\\
&= \sum_{\nu\in\mathcal{S}} q(\nu)\lambda_\nu\left(T_\mathbf{a}\right)\\
&= \langle T_\mathbf{a}\rangle_{\rho_{\text{in}}},
\end{aligned}$$

as required. We used Eq. (33) in line 2 above, Eq. (37) in line 4, Eq. (35) in line 5, and Eq. (33) in line 6. $\square$

VI. A QUBIT SCHEME OF QCSI WITH MATCHING WIGNER FUNCTION

A. Definition of the Wigner function

In this section we focus on the properties of a single Wigner function. We make the choice

$$\gamma(\mathbf{a}) = \mathbf{a}_Z \cdot \mathbf{a}_X \mod 4, \tag{39}$$

which has the important and rare consequence that the corresponding Wigner function factorizes, $W_{\rho\otimes\sigma} = W_\rho \cdot W_\sigma$ for all states ρ, σ. In fact, the factorization property already holds on the level of the Heisenberg-Weyl operators Eq. (5),

$$T_\mathbf{a} \otimes T_\mathbf{b} = T_{\mathbf{a}+\mathbf{b}}. \tag{40}$$

B. The set $\mathcal{O}$ of directly measurable observables

Lemma 10 *For $\gamma(\boldsymbol{a}) = \boldsymbol{a}_Z \cdot \boldsymbol{a}_X \mod 4$, the unique non-extendable consistent set $\mathcal{O}$ of directly measurable observables is the set of all local Pauli operators, $\mathcal{O} = \{\pm X_i, \pm Y_i, \pm Z_i |, i = 1, .., n\}$.*

This means first of all that W^0 *has* a corresponding QCSI scheme, and, perhaps surprisingly, spatial locality plays a role in it. Below, we first prove Lemma 10, and then flesh out the QCSI scheme corresponding to W^0.

Proof of Lemma 10. We first show that the set $\mathcal{O} = \{X_i, Y_i, Z_i |, i = 1, .., n\}$ satisfies the defining conditions

Eqs. (9), (15). Consider two commuting Pauli observables $T_{\mathbf{b}}$, $T_{\mathbf{c}}$ such that $T_{\mathbf{b}}$ is local to qubit k, and $T_{\mathbf{c}}$ is written as $T_{\mathbf{c}} = T_{\mathbf{c}'+\mathbf{c}''} = T_{\mathbf{c}'} \otimes T_{\mathbf{c}''}$, where $T_{\mathbf{c}'}$ acts non-trivially only on qubit k, and $T_{\mathbf{c}''}$ acts nontrivially only on the complement of qubit k. Then,

$$\begin{aligned} T_{\mathbf{b}} T_{\mathbf{c}} &= T_{\mathbf{b}} T_{\mathbf{c}'} \otimes T_{\mathbf{c}''} \\ &= i^{-\beta(\mathbf{b},\mathbf{c}')} T_{\mathbf{b}+\mathbf{c}'} \otimes T_{\mathbf{c}''} \\ &= i^{-\beta(\mathbf{b},\mathbf{c}')} T_{(\mathbf{b}+\mathbf{c}')+\mathbf{c}''} \\ &= i^{-\beta(\mathbf{b},\mathbf{c}')} T_{\mathbf{b}+\mathbf{c}}. \end{aligned}$$

Therein, in lines 1 and 3 we used the property Eq. (40).

Since $T_{\mathbf{b}}$ and $T_{\mathbf{c}}$ are commuting, $\beta(\mathbf{b}, \mathbf{c}') \in \{0, 2\}$. Since $T_{\mathbf{b}}$ is local, by going through all the cases of local Pauli operators we find that $\beta(\mathbf{b}, \mathbf{c}') \in \{0, \pm 1\}$. Thus, $\beta(\mathbf{b}, \mathbf{c}') = 0$ is the only choice that satisfies both constraints. Therefore,

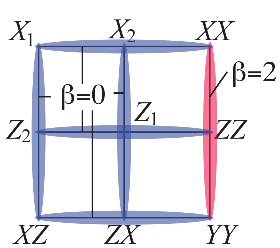
$$T_{\mathbf{b}+\mathbf{c}} = T_{\mathbf{b}} T_{\mathbf{c}}, \tag{41}$$

whenever $[T_{\mathbf{b}}, T_{\mathbf{c}}] = 0$, and $T_{\mathbf{b}}$ is local. Condition (9) is thus satisfied.

Next, since every multi-local Pauli operator can be written as a tensor product of local Pauli operators, and the local Pauli operators in such an expansion trivially commute, it follows that $V_M = V$, as required by condition (15). We have thus shown that $\{X_i, Y_i, Z_i |, i = 1, .., n\}$ is a possible set $\mathcal{O}$.

It remains to prove that the above $\mathcal{O}$ is non-extendable, i.e., that $\mathcal{O}$ cannot contain any additional observable without violating the condition Eq. (9). Uniqueness then follows from Lemma 4.

For $n = 1$ this is clear, and we only need to discuss the case of $n \geq 2$. To this end, consider the two-local Pauli operators, beginning with $Y \otimes Y$.



$Y \otimes Y$ is part of a context with $\beta = 2 \neq 0 \mod 4$. Therefore, with Lemma 1, $Y \otimes Y \notin \mathcal{O}$.

Now, conjugate the observables in the above diagram under a local Clifford unitary, and readjust the signs such that only observables $T_{\mathbf{a}}$ appear. In this way, any two-local Pauli observable can appear in the bottom left corner of the diagram. We observe that the four observables in the top left corner of the diagram will remain local under such a transformation. As we have shown, all local Pauli operators $T_{\mathbf{b}}$ satisfy Eq. (41) for all commuting $T_{\mathbf{c}}$. Hence, (i) The four βs involving local observables remain $\beta = 0$. (ii) The six β appearing in the square must sum to $2 \mod 4$, as per Mermin's argument. Combining these two facts, we find that the two β involving the observable in the bottom right corner of the diagram cannot simultaneously be zero. Hence this observable cannot be in $\mathcal{O}$. Thus, no two-local Pauli observable is in $\mathcal{O}$.

Next, consider a Pauli observable $T_{\mathbf{b}}$ with a support of size greater than 2. Be J a set of two sites in the support of $T_{\mathbf{b}}$, $J = \{j, k\} \subset \text{supp}(T_{\mathbf{b}})$, and denote by $T_{\mathbf{b}'}$ the restriction of $T_{\mathbf{b}}$ to J, and by $T_{\mathbf{b}''}$ the restriction of $T_{\mathbf{b}}$ to the complement of J. Then, with Eq. (40), $T_{\mathbf{b}} = T_{\mathbf{b}'+\mathbf{b}''} = T_{\mathbf{b}'} \otimes T_{\mathbf{b}''}$. Now consider a second Pauli operator $T_{\mathbf{c}}$ that commutes with $T_{\mathbf{b}}$ and has support on J only. Then, using the property Eq. (40),

$$\begin{aligned} T_{\mathbf{b}} T_{\mathbf{c}} &= T_{\mathbf{b}''} \otimes T_{\mathbf{b}'} T_{\mathbf{c}} \\ &= i^{\beta(\mathbf{b}',\mathbf{c})} T_{\mathbf{b}''} \otimes T_{\mathbf{b}'+\mathbf{c}} \\ &= i^{\beta(\mathbf{b}',\mathbf{c})} T_{\mathbf{b}''+(\mathbf{b}'+\mathbf{c})} \\ &= i^{\beta(\mathbf{b}',\mathbf{c})} T_{\mathbf{b}+\mathbf{c}}. \end{aligned}$$

By the previous argument for two-local operators, for any $T_{\mathbf{b}}$ with support on two or more qubits, a commuting two-local Pauli operator $T_{\mathbf{c}}$ can be found such that $\beta(\mathbf{b}', \mathbf{c}) = 2$. Then, with Lemma 1, $\pm T_{\mathbf{b}} \notin \mathcal{O}$. □

Remark: From the proof of Lemma 10 it is only a small step to show that the above $\mathcal{O}$ is the only consistent set for $\gamma(\mathbf{a}) = \mathbf{a}_Z \cdot \mathbf{a}_X \mod 4$, extendable or not. Namely, it follows from the condition (15) that all local Pauli operators must be present in $\mathcal{O}$. If any local Pauli operator is removed, then $V_M = V$ no longer holds.

From Eq. (4) it follows that the set Ω of free states are tensor products of one-qubit stabilizer states. The group of free unitary gates therefore is the local Clifford gates.

C. Magic states and universality

From the perspective of computational universality of QCSI, all we don't know yet is what the magic states are. Since all state preparations and measurements are local in the present situation, any entanglement needed in the computation must be brought in by the magic state. That is, there is only one big entangled magic state. Factors of tensor product states cannot be coupled by the free operations.

In fact, one possibility is to use as magic state a slightly modified cluster state. We consider a set of qubits located on the vertices of a square lattice graph. We denote the set of its sites by $\mathcal{V}$ and its adjacency matrix by Γ. We single out a subset $\mathcal{R} \subset \mathcal{V}$ of sites which are sufficiently sparse. Denote by A the observable $\frac{X+Y}{\sqrt{2}}$. With those definitions the resource state $|\Psi\rangle$ is the unique joint eigenstate with eigenvalue 1 of the stabilizer operators

$$K_a^X := X_a \bigotimes_{b \in \mathcal{V}} Z_b^{\Gamma_{ab}}, \text{ if } a \in \mathcal{V} \backslash \mathcal{R}, \tag{42}$$

$$K_a^A := A_a \bigotimes_{b \in \mathcal{V}} Z_b^{\Gamma_{ab}}, \text{ if } a \in \mathcal{R}. \tag{43}$$

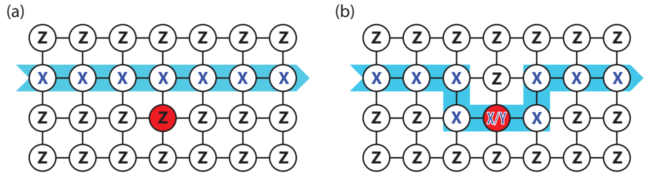


FIG. 5: QCSI with modified cluster state of Eqs. (42), (43) as magic state, which is subjected to measurements of local Pauli operators X_i, Y_j, Z_k, for all $i, j \neq i, k \neq i, j \in \mathcal{V}$. The role of the Z-measurements is to cut out of the plane a web corresponding to some layout of a quantum circuit, and the X-measurements drive the MBQC-simulation of this circuit [41]. The qubit in $\mathcal{R}$ is displayed in red. By "re-routing" a wire piece, one may choose between implementing and not implementing a non-Clifford gate. (a) Identity operation on the logical state space, (b) Logical gate $e^{i\pi/4\, Z}$.

That this leads to universal quantum computation is easily shown by standard arguments pertaining to measurement-based quantum computation (MBQC). See Fig. 5 for illustration.

While being a valid scheme of QCSI, this is also MBQC. The distinction between MBQC and QCSI is thus blurred. By various equivalence transformations, we can make the above computational scheme look more like the known QCSI schemes, or more like standard MBQC.

Equivalent scheme 1. In all QCSI schemes worked out to date [10], [9], [6], [7], the magic states are local to single or at most 2 particles. Although this is by no means necessary, we are used to those states being injected into the computation one by one. If desired, we may convert the above computational scheme into such a form, by conjugating it—the resource state, the measurable observables in $\mathcal{O}$, and the Wigner function W^0—under the unitary transformation

$$U_{\text{Ising}} = \prod_{i,j \in \mathcal{V}} \left(\Lambda(Z)_{i,j}\right)^{\Gamma_{ij}}. \qquad (44)$$

In this way, we arrive at the following equivalent computational scheme. The resource state $|\Psi\rangle$ is converted into a tensor product state of individual qubits being in the state $|+\rangle_i$, defined by $X|+\rangle = |+\rangle$, for $i \in \mathcal{V}\backslash\mathcal{R}$, and $|A\rangle_j$, defined by $A|A\rangle = |A\rangle$, for $j \in \mathcal{R}$. The new magic states are thus the local states $|A\rangle_j$.

The new set $\mathcal{O}_1$ of directly measure observables is $\mathcal{O}_1 = \{K_a^X, K_a^Y, Z_a,\, a = 1, .., n\}$, where $K_a^Y = Y_a \bigotimes_{b \in \mathcal{V}} {Z_b}^{\Gamma_{ab}}$.

Equivalent scheme 2. We note in Eq. (43) that stabilizer operators K^A of the magic state $|\Psi\rangle$ are not exactly stabilizer operators of cluster states. Therefore, we may apply the equivalence transformation

$$U_{\text{loc}} = \bigotimes_{j \in \mathcal{R}} e^{-i\pi/4\, Z_j},$$

and obtain as the new magic state the standard cluster state. The new measurable observables are

$$\mathcal{O}_2 = \{X_i, Y_i, A_j, A'_j, Z_k |\, a \in \mathcal{V}\backslash\mathcal{R}, j \in \mathcal{R}, k = 1..n\},$$

where $A' = (X - Y)/\sqrt{2}$. We note that the measurable observables which are not Zs are of the form

$$O_i = \cos\phi_i\, X_i \pm \sin\phi_i\, Y_i,$$

as standard in MBQC [41]. A minor deviation from the standard remains. Namely, for each site i, only a single setting out of two is available for the measurement angle, either $\phi_i = 0$ or $\phi_i = \pi/4$. In standard MBQC, any angle $\phi_i \in [0, \pi/2]$ may be chosen. However, the present restriction does not affect computational universality.

VII. CONCLUSION

In this paper, we have constructed a tomographically and informationally complete QCSI scheme for qubits in which contextality of the magic states is necessary for quantum computational universality. This case had remained open after the previously described schemes for qudits [6] and rebits [7]. More generally, we have shown that for all QCSI schemes on qubits satisfying the postulates (P1)-(P3), for $n \geq 2$ qubits, contextuality is necessary for quantum computational universality.

We also investigated the role of Wigner functions for the qubit case. The purpose of the Wigner function W^γ is to explain the near-classicality of the free sector of QCSI, i.e. the operations in QCSI which are not considered resources. We found the following features of interest:

(i) In contrast to the qudit and the rebit case described in the literature, for general QCSI schemes on qubits, the free unitary gates may introduce negativity into Wigner functions without compromising efficient classical simulability. The violation of positivity can be very large as measured by mana [38].

(ii) Any choice Eq. (5) of the phase convention γ for the Wigner function W^γ severely constrains the possible corresponding QCSI schemes. In particular, whenever a QCSI scheme exists for a given function γ, the non-extendable such scheme is unique.

An open problem is the classification of QCSI schemes, in particular the non-extendable ones. Enumeration by the different phase conventions γ for the Pauli operators Eq. (5) is not a satisfactory approach, since for the overwhelming majority of functions γ no corresponding QCSI scheme exists. A further open question is the characterization of the magic states for all QCSI schemes.

In this work, we have used classical simulation of quantum computation as a tool to identify quantum resources required for a speed-up. More generally, determining the cost of simulating universal quantum computing classically is fascinating open question; See e.g. [42], [43].

Acknowledgments. RR thanks Terry Rudolph for discussion. This work has been funded by NSERC (ND, CO and RR), SIQS (JBV), AQuS (JBV), Cifar (ND and RR), and IARPA (RR). RR is Fellow of the Cifar Quantum Information Program.

[1] J.S. Bell, *On the Problem of Hidden Variables in Quantum Mechanics*, Rev. Mod. Phys. **38**, 447 (1966).
[2] S. Kochen and E.P. Specker, *The Problem of Hidden Variables in Quantum Mechanics*, J. Math. Mech. **17**, 59 (1967).
[3] N. D. Mermin, *Hidden Variables and the Two Theorems of John Bell*, Rev. Mod. Phys. **65**, 803 (1993).
[4] Samson Abramsky, Adam Brandenburger, *The sheaf-theoretic structure of non-locality and contextuality*, New J. Phys. **13**, 113036 (2011).
[5] A. Acín, T. Fritz, A. Leverrier, A. Belén Sainz, *A Combinatorial Approach to Nonlocality and Contextuality*, arXiv:1212.4084.
[6] M. Howard *et al.*, Nature (London) **510**, 351 (2014).
[7] N. Delfosse, P. Allard-Guerin, J. Bian and R. Raussendorf, *Wigner Function Negativity and Contextuality in Quantum Computation on Rebits*, Phys. Rev. **X** 5, 021003 (2015).
[8] S. Bravyi and A. Kitaev, *Universal Quantum Computation with Ideal Clifford Gates and Noisy Ancillas*, Phys. Rev. A **71**, 022316 (2005).
[9] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, *Negative Quasi-probability as a Resource for Quantum Computation*, New J. Phys. **14**, 113011 (2012).
[10] E. F. Galvao, *Discrete Wigner Functions and Quantum Computational Speedup*, Phys. Rev. A **71**, 042302 (2005).
[11] Cecilia Cormick, Ernesto F. Galvao, Daniel Gottesman, Juan Pablo Paz, Arthur O. Pittenger, *Classicality in discrete Wigner functions*, Phys. Rev. A **73**, 012301 (2006).
[12] E. Wigner, *On the Quantum Correction For Thermodynamic Equilibrium*, Phys. Rev. **40**, 749 (1932).
[13] K.S. Gibbons, M.J. Hoffman, and W.K. Wootters, *Discrete Phase Space Based on Finite Fields*, Phys. Rev. A **70**, 062101 (2004).
[14] D. Gross, *Computational power of quantum many-body states and some results on discrete phase spaces*, PhD Thesis, Imperial College London, 2005.
[15] A. Mari, J. Eisert, *Positive Wigner functions render classical simulation of quantum computation efficient*, Phys. Rev. Lett. **109**, 230503 (2012).
[16] R.L. Hudson, *When is the Wigner quasi-probability density non-negative?*, Rep. Math. Phys. 6, 249 (1974).
[17] Robert W. Spekkens, *Negativity and contextuality are equivalent notions of nonclassicality*, Phys. Rev. Lett. **101**, 020401 (2008).
[18] Huangjun Zhu, *Permutation symmetry determines the discrete Wigner function*, arXiv:1504.03773.
[19] Huangjun Zhu, *Multiqubit Clifford groups are unitary 3-designs*, arXiv:1510.02619.
[20] N. D. Mermin, *Extreme quantum entanglement in a superposition of macroscopically distinct states*, Phys. Rev. Lett. **65**, 1838 (1990) .
[21] V. Scarani, A. Acin, E. Schenck, and M. Aspelmeyer, *Nonlocality of cluster states of qubits*, Phys. Rev. A **71**, 042325 (2005).
[22] O. Gühne, G. Toth, P. Hyllus, and H. J. Briegel, *Bell inequalities for graph states*, Phys. Rev. Lett. **95**, 120405 (2005).
[23] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis, California Institute of Technology, 1997.
[24] Mark Howard, Eoin Brennan, Jiri Vala, *Quantum Contextuality with Stabilizer States*, Entropy **15**, 2340 (2013).
[25] Y. Han, L. Hemaspaandra and T. Thierauf, *Threshold computation and cryptographic security*, SIAM J. on Computing **26**, 59 (1997).
[26] S. Aaronson, *Quantum computing, post-selection and probabilistic polynomial time*, Proc. R. Soc. A **461**, 3473 (2005).
[27] S. Aaronson and A. Arkhipov, Proc. of ACM STOC, 333-342, (2011).
[28] M. J. Bremner, R. Jozsa, and D. J. Shepherd, *Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy*, Proc. R. Soc. A **467**, 459 (2011).
[29] Tomoyuki Morimae, Keisuke Fujii, Joseph F. Fitzsimons, *On the hardness of classically simulating the one clean qubit model*, Phys. Rev. Lett. **112**, 130502 (2014).
[30] Matty J. Hoban *et al.*, *Measurement-based classical computation*, Phys. Rev. Lett. **112**, 140505 (2014).
[31] A. R. Calderbank, Peter W. Shor, *Good Quantum Error-Correcting Codes Exist*, Phys. Rev. A **54**, 1098 (1996).
[32] Andrew Steane, *Multiple Particle Interference and Quantum Error Correction*, Proc.Roy.Soc.Lond. A **452**, 2551 (1996).
[33] D. W. Kribs, R. Laflamme, D. Poulin, and M. Lesosky, *Operator quantum error correction*, QIC **6**, 382 (2006).
[34] D. Poulin, *Stabilizer formalism for operator quantum error correction*, Phys. Rev. Lett. **95**, 230504 (2005).
[35] D. Bacon, *Operator quantum error-correcting subsystems for self-correcting quantum memories*, Phys. Rev. A **73**, 012340 (2006).
[36] P. W. Shor, *Scheme for reducing decoherence in quantum computer memory*, Phys. Rev. A **52**, 2493 (1995).
[37] H. Bombin, *Topological Subsystem Codes*, Rev. A **81**, 032301 (2010).
[38] Victor Veitch, Seyed Ali Hamed Mousavian, Daniel Gottesman, Joseph Emerson, *The Resource Theory of Stabilizer Computation*, New J. Phys. **16**, 013009 (2014).
[39] Hakop Pashayan, Joel J. Wallman, Stephen D. Bartlett, *Estimating outcome probabilities of quantum circuits using quasiprobabilities*, Phys. Rev. Lett. **115**, 070501 (2015).
[40] Joel J. Wallman, Stephen D. Bartlett, *Nonnegative subtheories and quasiprobability representations of qubits*, Phys. Rev. A **85**, 062121 (2012).
[41] R. Raussendorf and H.J. Briegel, *A one-way quantum computer*, Phys. Rev Lett **86**, 5188 (2001).
[42] M. Van den Nest, W. Dür, G. Vidal, H. J. Briegel, *Classical simulation versus universality in measurement based quantum computation*, Phys. Rev. A **75**, 012337 (2007).
[43] Sergey Bravyi, Graeme Smith, John Smolin, *Trading classical and quantum computational resources*, arXiv:1506.01396.